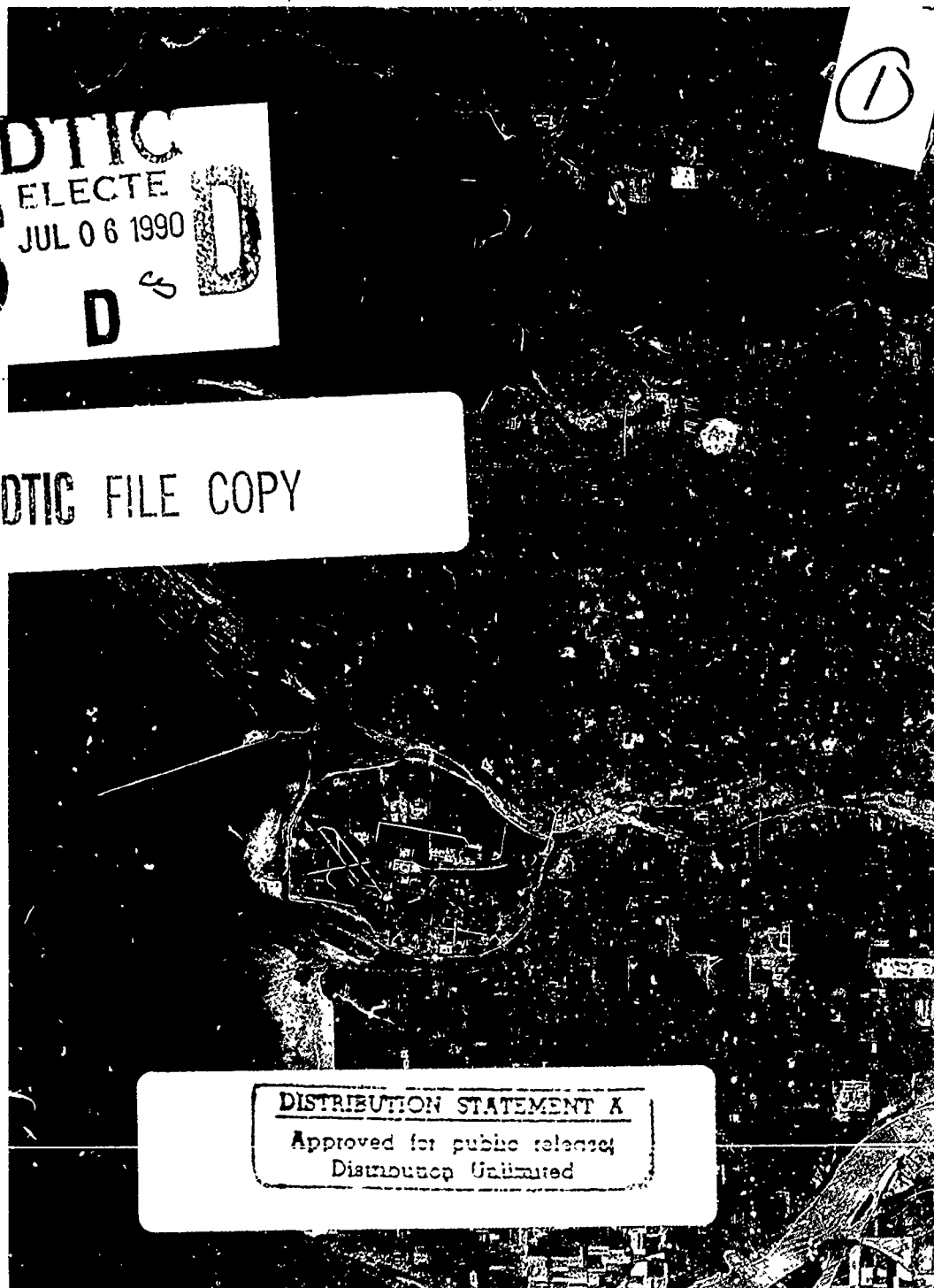


Quantitative Remote Sensing: An Economic Tool for the Nineties

N00014-89-J-1969



DTIC
ELECTE
JUL 06 1990

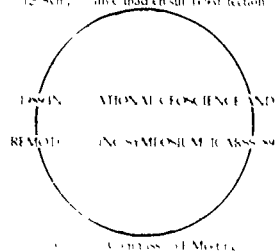
DTIC FILE COPY

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Volume 2. Tuesday July 11

90 07 3 229



VANCOUVER, CANADA JULY 10-14, 1989

AD-A223 843

VOLUME 2

IEEE No. 89CH2768-0
LIBRARY OF CONGRESS No. 89-84217

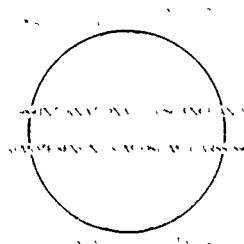
IGARSS '89

12th CANADIAN SYMPOSIUM ON REMOTE SENSING
12e SYMPOSIUM CANADIEN SUR LA TELEDETECTION

REMOTE SENSING:

AN ECONOMIC TOOL FOR THE NINETIES

JULY 10 - 14 1989 VANCOUVER, CANADA



SPONSORS:
IEEE, CRSS, URSI

AVAILABLE FOR \$186 FOR SET per Institute
of Electrical and Electronics Engineers,
Inc. 345 East 47th St., New York, NY
10017
TELECON 7/5/90 VG

| | |
|----------------------------|-------------------------------------|
| Acquisition For | |
| NTIS CRASI | <input checked="" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By <i>\$186.00 per set</i> | |
| Vols <i>1-5</i> per cell | |
| Availability Codes | |
| Dist | Avail and/or Special |
| <i>A-1</i> | <i>21</i> |



**IGARSS '89
REMOTE SENSING: AN ECONOMIC
TOOL FOR THE NINETIES
VANCOUVER, CANADA
JULY 10-14, 1989**

COPYRIGHT - 1989 by IEEE

**Proceedings published by:
IGARSS '89 12th Canadian Symposium on Remote
Sensing**

Printed in Canada

IGARSS '89 - 12th CANADIAN SYMPOSIUM ON REMOTE SENSING
IGARSS '89 - 12th SYMPOSIUM CANADIEN SUR LA TELEDETECTION

IGARSS '89 - 12th CANADIAN SYMPOSIUM
ON REMOTE SENSING SPONSORS
SPONSORS COMMANDITAIRES

IGARSS '89 - 12th Canadian Symposium on Remote Sensing is sponsored by the Canadian Remote Sensing Society, the IEEE Geoscience and Remote Sensing Society, the International Union of Radio Science, and the Canada Centre for Remote Sensing.

It is co-sponsored by:

B.C. Ministry of Environment and Parks
B.C. Ministry of Forests, Inventory Branch
B.C. Research Corporation
Binary Image Corporation
Canadian Forest Service
Cominco Ltd.
Fisheries and Oceans Canada
MacDonald, Dettwiler and Associates Ltd.
National Research Council Canada
Pamap Graphics Ltd.
University of British Columbia

IGARSS '89 - 12th CANADIAN SYMPOSIUM ON
REMOTE SENSING ORGANIZING COMMITTEE

Dr. John S. MacDonald, General Chairman

Dr. Ed Jull, Vice Chairman, URSI
Dr. Peter Murtha, Vice Chairman, CRSS
Dr. Keith Raney, Vice Chairman, IEEE/GRS-S

Dr. Jim Gower, Technical Program Chairman

Dr. Josef Cihlar, Technical Co-Chairman, CRSS
Dr. David Goodenough, Technical Co-Chairman, IEEE/GRS-S

M. Jean-Marie Dubois, French Transcription
Mr. Frank Hegyi, Social Program and Tours
Ms. Nedenia Holm, Secretary
Dr. Y. Jim Lee, Exhibits
Mr. Peter Louis, Finance
Dr. Pam Sallaway
Dr. Hans Schreier, UBC Facilities/Registration
Dr. Mark Sondheim, Publicity
Dr. Mit Tilkov
Dr. Geoff Tomlins, Publication

CRSS Technical Committee

Chair: J. Cihlar, Canada Centre for Remote Sensing, Canada
F.J. Ahern, Canada Centre for Remote Sensing, Canada
Kohei Arai, National Space Development Agency of Japan, Japan
F. Bonn, Université de Sherbrooke, Canada
G.A. Borstad, G.A. Borstad Associates Ltd., Canada
J. Brown, Canada Centre for Remote Sensing, Canada
R. Currie, Pacific Geoscience Centre, Canada
G. Duchossois, European Space Agency, Canada
A. Fabbri, Canada Centre for Remote Sensing, Canada
B.E. Goodison, Atmospheric Environment Service, Canada
B. Guidon, Canada Centre for Remote Sensing, Canada
F.G. Hall, NASA Goddard Space Flight Centre, USA
D. Halpern, Jet Propulsion Laboratories, USA
F. Hegyi, Ministry of Forests, Canada
J. Iisaka, Canada Centre for Remote Sensing, Canada
J. Kosalos, Honeywell Marine Systems, USA
L. McNutt, Canada Centre for Remote Sensing, Canada
J.B. Mercer, Intera Technologies Ltd., Canada
R.E. Murphy, NASA/Headquarters, USA
R.A. O'Neil, Canada Centre for Remote Sensing, Canada
P. Sallaway, PAMAP Graphics Ltd., Canada
L. Sayn-Wittgenstein, Canada Centre for Remote Sensing, Canada
R. Sianey, Radarsat Project Office, Canada
W.M. Ström, PCI Inc., Canada
T. Toutin, Canada Centre for Remote Sensing, Canada
B. Woodham, University of British Columbia, Canada
V. Zsilinszky, Ontario Centre for Remote Sensing, Canada

IEEE/GRSS Technical Committee

Chair: David G. Goodenough, Canada Centre for Remote Sensing, Canada
R.C. Beal, The John Hopkins University, USA
G. Brachet, SPOT Image, France
F. Carsey, Jet Propulsion Laboratories, USA
K.R. Carver, University of Massachusetts, USA
R. Cheney, National Ocean Service NOAA, USA
J. Cimino, Jet Propulsion Laboratories, USA
M. Ehlers, University of Maine, USA
A.F.H. Goetz, University of Colorado, USA
K.I. Itten, University of Zurich-Irchel, Switzerland
E. Ledrew, University of Waterloo, Canada
C.A. Luther, Office of Naval Research, USA
J.P. Muller, University College London, UK
J.A. MacDonald, MacDonald Dettwiler Associates, Canada
J. Megier, Ispra Establishment, Italy
R.K. Raney, Radarsat Project Office, Canada
P.N. Slater, University of Arizona, USA
C.T. Swift, University of Massachusetts, USA
Mikio Takagi, University of Tokyo, Japan
G.F. West, University of Toronto, Canada

URSI Technical Committee

Chair: J.F.R. Gower, Institute of Ocean Sciences, Canada
J.R. Apel, The John Hopkins University, USA
R.K. Crane, Dartmouth College, USA
J.L. Green, NOAA Aeronomy Laboratory, USA
A. Guissard, Université Catholique de Louvain, Belgium
D.C. Hogg, University of Colorado, USA
R. Humphreys, MacDonald Dettwiler Associates, Canada
E. Jull, University of British Columbia, Canada
R.K. Moore, University of Kansas Centre for Research Inc., USA
D.W. Oldenburg, University of British Columbia, Canada
R. Olsen, Communications Research Centre, Canada
B. Segal, Communications Research Centre, Canada
O.H. Shemdin, Ocean Research and Engineering, USA
J.R. Wait, University of Arizona, USA
J. Walsh, Memorial University of Newfoundland, Canada

Book 2
Table of Contents

| | Page |
|--|------|
| Comparison of theoretical and measured brightness temperature of a soybean canopy R.H. Lang, D.M. LeVine, P. O'Neill, S. Saatchi, NASA/GSFC, Greenbelt, MD, USA, O. Yazici, George Washington Univ., Washington, DC, USA, T. Jackson, US Dept. of Agric., Beltsville, MD, USA | 400 |
| Ku-Band polarization characteristics of wheat and canola M.R. Hinds, A.G. Wacker, G.J. Sofko, M.P. Poettcker, Univ. of Saskatchewan, Saskatoon, Sask., Canada | 405 |
| A microwave backscatter experiment of a burdock field by an X-Band FM-CW radar T. Kurosu, T. Sultz, T. Umehara, Min. of Posts and Telecommun., Tokyo, Japan | 409 |
| Radar en agriculture: resultats preliminaires sur un site agricole au Quebec K.P.B. Thomson, R. Landry, G. Edwards, M. Cantin, Univ. Laval, Ste-Foy, Que., Canada, H. Gwyn, Univ. de Sherbrooke, Sherbrooke, Que., Canada | 415 |
| Inventaires agricoles par teledetection: Validite des estimateurs bases sur la matrice de confusion J. Delince, Inst. de Teledetection, CCR, Ispra, Italy, G. Gerard, UCL, Louvain-la-neuve, Belgium | 418 |
| Supervised image interpretation of SPOT-data for the detection of agricultural parcels in different vegetation types in Zaire (Bandundu) L. Daels, R. Goossens, T. Ongena, M. De Caluwe, Rijks Univ., Gent, Netherlands | 419 |
| Quantifying forage and corn production with image analysis and GIS techniques H. Schreier, S. Brown, A. Ryan, R. Wiart, Univ. of British Columbia, Vancouver, BC, Canada | 421 |
| The effects of free canopy water on SAR crop separability B. Brisco, R.J. Brown, CCRS, Ottawa, Ont., Canada, T.J. Pultz, INTERA Tech. Ltd., Ottawa, Ont., Canada | 424 |
| Analysis of C-Band SAR data in an agricultural context N. Prout, J. Gairns, C. Hutton, D. Mullins, INTERA Tech. Ltd., Ottawa, Ont., Canada, A. Mack, Agric. Canada, Ottawa, Ont., Canada | 429 |
| Preliminary evaluation of multi-date SAR data for the identification of agricultural crops in southern Ontario J. Fischer, R. Mussakowski, Ontario Centre for Rem. Sens., North York, Ont., Canada | 430 |
| Recognition of three-dimensional objects from single images D.G. Lowe, Univ. of British Columbia, Vancouver, BC, Canada | 434 |
| Frame camera and image contextual constraints for stereo matching in aerial scenes; D.W. Paglieroni, Ford Aerospace Corp., San Jose, CA, USA | 435 |
| Geometric transformation and terrain correction of remotely-sensed imagery J.T. Parr, Analytic Sci. Corp., Reading, MA, USA | 440 |

(to piii)

| | Page |
|---|------|
| Une methode de detection d'un reseau hydrographique a partir d'un modele numerique de terrain P. Guillotel, Ecole Nat. Sup. des Telecom. de Bretagne, Brest, France, R. Simard, B. Goze, DIGIM, Montreal, Que., Canada, C. Roux, Ecole Nat. Sup. des Telecom. de Bretagne France | 442 |
| Error correction of digital elevation models produced by automatic matching of digital stereo images J.A. Ostrowski, D.C. He, Horler Info. Inc., Ottawa, Ont., Canada | 446 |
| Least squares prediction using on-board data in bundle adjustment for SPOT imagery L.-C. Chen, L.-H. Lee, Nat. Cen. Univ., Chung-Li, Taiwan, Rep. China | 450 |
| Investigation of slope and aspect effects on image grey values in geocoded SAR images T. Bayer, R. Winter, G. Schreier, DFVLR, German Aerospace Res. Est., Oberpfaffenhofen, FR Germany | 454 |
| Quantitative analysis and classification of SeaMARC II 12 kHz side-scan sonar data T. Reed, Hawaii Inst. of Geophysics, Honolulu, HI, USA | 458 |
| Le programme SPOT: planification a moyen et long terme sur la base des besoins du marche G. Brachet, SPOT IMAGE, Toulouse, France | 462 |
| Evaluation of SPOT HRV image data received in Japan N. Fujimoto, Y. Takahashi, T. Moriyama, M. Shimada, H. Wakabayashi, NASDA, EOC, Saitama-ken, Japan, Y. Nakatani, Rem. Sens. Tech. Cen., Japan, S. Obayashi, Science Univ. of Tokyo, Japan | 463 |
| Three-dimensional analysis of SPOT HRV data M. Mori, Kinki Univ., Iizuka, Japan, K. Gotoh, Nagasaki Univ., Nagasaki, Japan | 467 |
| Removal of the topographic effect on SPOT-1 HRV multispectral imagery in mountainous terrain C. Yang, A. Vidal, Lab. Commun de Teledetec., CEMAGREF EMGREF, Montpellier, France | 471 |
| Computer assisted land cover mapping with SPOT in Indonesia J.-P. Gastellu-Etcheberry, Gadjah Mada Univ., Yogyakarta, Indonesia, D. Ducros-Gambert, Paul Sabatier Univ., Toulouse, France | 475 |
| Optimisation de la restitution automatique du relief a partir d'images SPOT A. Jaton, R. Simard, A. Leclerc, S.R. Haja, Digim, Inc., Montreal, Que., Canada | 476 |
| Approche multitemporelle et texturale des donnees SPOT multispectrales pour le suivi agricole en Afrique Centrale P. Jacques, M. Massart, J. Wilmet, Laboratoire de Teledetection, Louvain la Neuve Belgique | 480 |
| Applications of stereo SPOT for mapping J-P Muller, I.J. Dowman, University College London, London U.K. | 484 |
| Radiometric comparison of Landsat-5 TM and SPOT HRV 1 sensors for the use in multiple sensor approaches J. Hill, D. Aifadopoulou, Comm. for the Eur. Communities, Ispra, Italy | 485 |
| Neural network approaches versus statistical methods in classification of multisource remote sensing data J.A. Benediktsson, P.H. Swain, O.K. Ersoy, Purdue Univ., W. Lafayette, IN, USA | 489 |

| | Page |
|---|------|
| Statistical methods for cover class mapping using remotely sensed data N.A. Campbell, J.F. Wallace, CSIRO, Wembley, W.A., Australia | 493 |
| Spatial-temporal autocorrelated model for contextual classification of satellite imagery N. Khazenie, M.M. Crawford, Univ. of Texas at Austin, Austin, TX, USA | 497 |
| Eigenvector transformation and dimension size reduction in remote sensing data processing B. Li, W.M. Moon, Univ. of Manitoba, Winnipeg, Man., Canada | 503 |
| A method for multi-dimensional image segmentation K. Hartt, M.J. Carlotto, M.W. Brennen, Analytic Sci. Corp., Reading, MA, USA | 509 |
| An adaptive reconstruction system for spatially correlated multi-spectral, multi-temporal images S. Lee, M.M. Crawford, Univ. of Texas, Austin, TX, USA | 513 |
| A method of statistical multisource classification with a mechanism to weight the influence of the data sources J.A. Benediktsson, P.H. Swain, Purdue Univ., W. Lafayette, IN, USA | 517 |
| Observations of clouds and fog with a 1.4 mm wavelength radar J.B. Mead, R.E. McIntosh, Univ. of Massachusetts, Amherst, MA, USA | 521 |
| 95 GHz short pulse radar E.A. Uliana, B.S. Yapple, US Naval Res. Lab., Washington, DC, USA, N.C. Chu, M. Newkirk, Bendix Field Engin., Oxon Hill, MD, USA | 524 |
| 140 GHz scatterometer measurements T.F. Haddock, F.T. Ulaby, J. East, Univ. of Michigan, Ann Arbor, MI, USA | 525 |
| The CRL millimeter wave scatterometers K. Okamoto, T. Ihara, J. Awaka, H. Yamasaki, A. Takahashi, Min. of Posts & Telecomm., Tokyo, Japan | 529 |
| Millimeter-wave propagation measurements through conifer vegetation D.L. Jones, R.H. Espeland, Inst. for Telecomm. Sci., Boulder, CO, USA | 533 |
| Millimeter wave scattering model for a leaf K. Sarabandi, F.T. Ulaby, T.B.A. Senior, Univ. of Michigan, Ann Arbor, MI, USA | 534 |
| Measurements and modeling of millimeter-wave scattering from tree canopies F.T. Ulaby, T.F. Haddock, Y. Kuga, Univ. of Michigan, Ann Arbor, MI, USA | 538 |
| Spaceborne technology contributions to hydrological studies in the context of global change V.V. Salomonson, NASA/GSFC, Greenbelt, MD, USA | 539 |
| Remote sensing and hydrology - the commercial prospects J.W. Trevett, ICRSDT, Hunting Technical Services Ltd, Hemel Hempstead UK | 542 |
| Cost-effectiveness of Landsat TM classification by operations staff S.-L.R. Konrad, R.S. Rempel, Ducks UnLtd. Canada, Winnipeg, Man., Canada | 543 |
| Digital image processing of multitemporal Landsat data and its applications in ground water exploration C.S.S. Reddy, D.J. Campagna, D.W. Levandowski, Purdue Univ., W. Lafayette, IN, USA | 544 |

| | Page |
|--|------|
| Combining remote sensing and conventional data for assessing water resources in Northern Thailand C. Saowaporn, Chiang Mai Univ., Chiang Mai, Thailand, S.I. Solomon, E.D. Soulis, Univ. of Waterloo, Waterloo, Ont., Canada | 548 |
| Using NOAA data for hydrological modeling G. Kite, Canada Climate Centre, Saskatoon, Sask., Canada | 553 |
| Snow cover monitoring by satellites and real time runoff forecasts K. Seidel, U. Burkart, R. Baumann, Inst. fur Kommun. ETHZ, Zurich, Switzerland, J. Martinec, Eidg. Inst. fur Schnee- & Lawin., Davos Switzerland | 558 |
| Residual snow patch mapping in Arctic Canada using Landsat TM images J.T. Gray, Univ. of Montreal, Montreal, Que., Canada, B. Lauriol, Univ. d'Ottawa, Ottawa, Ont., Canada, D. Bruneau, P. Briand, Univ. de Montreal, Montreal, Qc, Canada | 562 |
| The technological challenge of future space radar systems H.M. Braun, G.H. Rausch, H. Kappel, Dornier Sys. GMBH, Friedrichshafen, FR Germany | 566 |
| Multimission capability of the European Polar Platform R. Benz, G. Gebauer, F. Tanner, Dornier GmbH, Friedrichshafen, FR Germany, E. Zeiss, ESTEC, Noordwijk, The Netherlands | 570 |
| Upgrades to the Australian Centre for Remote Sensing - A total ground station solution J. Friedel, MacDonald Dettwiler, Richmond, BC, Canada, C. McMaster, Austr. Cen. for Rem. Sens., Belconnen, Australia | 575 |
| Ecuador ground receiving station - component subsystems K. Cox, F. Pleasance, MacDonald Dettwiler, Richmond, BC, Canada | 579 |
| Aristoteles - Description of the earth gravity field recovery mission R. Benz, M. Langemann, M. Gramolla, Dornier GmbH, Friedrichshafen, FR Germany, G. Mecke, ESTEC, Noordwijk, The Netherlands | 582 |
| Two way satellite communication for telemetrology and remote control H. Hanebrekke, Norwegian Inst. of Tech., Trondheim, Norway | 587 |
| Influence of the relief on radiometry and resolution of a SAR image Cl. Jung, D. Vidal Madjar, CNET/CRPE, Issy-Les-Moulineaux, France, Ph. Rebillard, Sepimage, Puteaux, France | 591 |
| Multiple-image SAR shape from shading J. Thomas, W. Kober, F. Leberl, VEXCEL Corp., Boulder, CO, USA | 592 |
| Development of a shape-from-shading technique for the extraction of topographic models from individual spaceborne SAR images B. Guindon, CCRS, Ottawa, Ont., Canada | 597 |
| A method of determining the pixel geolocation of SAR images in oceanic and polar regions U. Tretter, DFVLR, Oberpfaffenhofen, FR Germany | 603 |
| Prototype SAR geocoding algorithms for ERS-1 and SIR-C / X-SAR images A. Roth, H. Craubner, Th. Bayer, DFVLR, Oberpfaffenhofen, FR Germany | 604 |
| Automated ice motion tracking at the Alaska SAR facility R. Kwok, NASA/JPL, Pasadena, CA, USA | 608 |

| | Page |
|---|------|
| Interpolation and sampling in SAR images S. Quegan, Univ. of Sheffield, Sheffield, UK | 612 |
| Minimum mean-square error sampling for remotely sensed imagery G.W. Wornell, D. Kauffman, B. Sharpe, MacDonald Dettwiler, Richmond, BC, Canada | 616 |
| Effect of resampling on the geometric and radiometric fidelity of digital images E. Derenyi, R.K. Saleh, Univ. of New Brunswick, Fredericton, NB, Canada | 620 |
| Effects of the ephemeris error on effective pointing for a spaceborne SAR; <i>→ lto p vii)</i> M.Y. Jin, NASA/JPL, Pasadena, CA, USA | 624 |
| Statistically based unsupervised hierarchical image segmentation algorithm with a blurring corrector S. Lee, M.M. Crawford, Univ. of Texas at Austin, Austin, TX, USA | 630 |
| Pixel relaxation labelling using a diminishing neighbourhood effect T. Lee, Chinese Univ. of Hong Kong, Shatin, NT, Hong Kong, J.A. Richards, Univ. New South Wales, Campbell, Australia | 634 |
| Spectral matching in supervised classification of terrain features Z. Ma, C.E. Olson Jr., Univ. of Michigan, Ann Arbor, MI, USA | 638 |
| Improved classification using imagery from multiple satellites D. Kauffman, B.C. Robertson, MacDonald Dettwiler, Richmond, BC, Canada | 642 |
| Classification of satellite images using contextual classifiers F.G. Alonso, S. Lopez Soria, Inst. Nacional de Investig. Agrarias, Madrid, Spain | 645 |
| Classification of SPOT image using spectral and spatial features of primitive regions with nearly uniform color T. Kusaka, H. Egawa, Y. Kawata, Kanazawa Inst. of Tech., Ishikawa, Japan | 649 |
| High resolution satellite image texture for moderate relief terrain analysis D.R. Peddle, NORDCO Ltd., St. John's, Nfld., Canada, S.E. Franklin, Univ. of Calgary, Calgary, Alta., Canada | 653 |
| Integration of context classifiers with GIS D. Booth, T.R.E. Chidley, W.G. Collins, Aston Univ., Birmingham, UK | 656 |
| Changes in canopy and branch reflectance along an SO ₂ gradient on Northern Vancouver Island B. Curtiss, Univ. of Colorado, Boulder, CO, USA, S.L. Ustin, Univ. of California, Davis, CA, USA | 660 |
| Laboratory fluorescence, reflectance and chlorophyll measurements of foliage from a stressed spruce forest C. Banninger, Inst. for Im. Proc. and Comp. Graphics, Graz, Austria | 661 |
| Reflectance and transmittance of snow at high spectral resolution J. Dozier, NASA/JPL, Pasadena, CA, USA, R.E. Davis, Univ. of California, Mammoth Lakes, CA, USA, A.W. Nolin, Univ. of California, Santa Barbara, CA, USA | 662 |
| Reflectance spectra of vegetation growing on mine sites in the Canadian Shield V. Singhroy, F. Kenny, Min. of Natural Res., North York, Ont., Canada, J. Speinger, Ont. Geolog. Surv., Sudbury, Ont., Canada | 665 |

| | Page |
|--|------|
| Bidirectional reflectances of three soil surfaces and their characterization through model inversion D.W. Deering, NASA/GSFC, Greenbelt, MD, USA, T.F. Eck, ST Systems Corp., Lanham, MD, USA, J. Otterman, Tel Aviv Univ., Ramat Aviv, Israel | 670 |
| The various chlorophyll fluorescence signatures as a basis for physiological ground truth control in remote sensing of forest decline U. Rinderle, H.K. Lichtenthaler, Univ. of Karlsruhe, Karlsruhe, FR Germany | 674 |
| Sensitivity of reflected radiation to organic matter content of typical Ontario soils A. Nurwadjadi, R. Protz, Univ. of Guelph, Guelph, Ont., Canada, J. Cihler, CCRS, Ottawa, Ont., Canada | 678 |
| Spectral components analysis equations and their application to rice C.L. Wiegand, Agric. Res., USDA, Weslaco, TX, USA, M. Shibayama, Y. Yamagata, Nat. Inst of Agro-Env. Sci., Ibaraki, Japan | 679 |
| Utilization des mesures radiometriques de terrain pour la cartographie satellitaire des sols d'une region semi-aride du Maroc M. Hinse, Q.H.J. Gwyn, Univ. de Sherbrooke, Sherbrooke, Que., Canada, A. Merzouk, Depart. des Sci. du Sol, Rabat, Maroc, F. Bonn, Univ. de Sherbrooke, Sherbrooke, Qc, Canada | 684 |
| Basis Frequency Analysis, BAFRAN, - towards a fresh approach to analyse the spectrum of a short record M.P. Beddoes, Univ. of British Columbia. Vancouver, BC, Canada | 688 |
| Convergent fronts in a partially mixed estuary J.M. Anderson, H. Evans, D. Djavadi, Univ. of Dundee, Dundee UK | 691 |
| Classification of Puget Sound nearshore habitats using aircraft multispectral scanner imagery F. Mynar, R.D. Weerackoon, M.V. Olsen, Lockheed Eng. and Sci. Co., Las Vegas, NV, USA, R.S. Lunetta, US Env. Prot. Agency, Las Vegas, NV, USA | 695 |
| Use of thematic mapper data for mapping of water quality K. Sorensen, Norwegian Inst. for Water Res., Blindern, Norway, J. Nilsen, Norwegian Hydro. Lab, Trondheim, Norway, H.V. Sebo, E. Holbek-Hanssen, Norwegian Comput. Cen., Blindern, Norway | 696 |
| A data-fusion approach to sea water monitoring D.D. Giusto, L. Parodi, G. Vernazza, Univ. of Genoa, Genoa, Italy | 697 |
| Utility of airborne electromagnetic profiler data for determining water depths and characterizing coastal sediments and marsh soils R.E. Pelletier, S.T. Wu, Stennis Sp. Cen., MS, USA | 701 |
| Bathymetry retrieval algorithm for coastal water [case II] from Landsat - 5 Thematic Mapper Data A.K. Mishra, P.N. Sridhar, K.S. Prasad, Nat. Rem. Sens. Agency, Hyderabad, India | 705 |
| Normalization of tidal fluctuations for use in satellite remote bathymetry Wei Ji, D.L. Civco, W.C. Kennard, Univ. of Connecticut, Storrs, CT, USA | 709 |
| A study into the responses of the NOAA-n AVHRR reflective channels over water targets S.C. Gallegos, T.I. Gray, M.M. Crawford, Univ. of Texas at Austin, Austin, TX, USA | 712 |
| Overview and highlights from the NORCSEX'88: a prelaunch ERS-1 experiment J.A. Johannessen, O.M. Johannessen, Nansen Rem. Sens. Cen., Bergen, Norway, R.A. Shuchman, Env. Res. Inst. of Michigan, Ann Arbor, USA | 716 |

| | Page |
|--|------------|
| 17 SAR detection of mesoscale ocean circulation features as during NORCSEX '88 ; R.A. Shuchman, D.R. Lyzenga, Env. Sci. Inst. of Michigan, Ann Arbor, MI, USA, J.A. Johannessen, O. Johannessen, Nansen Res. Sens. Centre, Bergen, Norway, C. Livingstone, CCRS, Ottawa, Ont., Canada | 720 |
| Surface current measurements at ocean fronts D.R. Johnson, F.M. Fetterer, Stennis Space Center, Stennis, MS, USA | 726 |
| Surface currents during NORCSEX'88, as measured by a land- and a ship-based HF-radar H.-H. Essen, Univ. Hamburg, Hamburg, FR Germany | 730 |
| Mesoscale variability in the Haltenbanken region mapped by Geosat topography data L.H. Pettersson, J.A. Johannessen, T.A. Johansen, P. Samuel, Nansen Rem. Sens. Cen., Bergen, Norway | 734 |
| Mesoscale variability during NORCSEX '88 derived from AVHRR surface structures, ADCP currents and a numerical model P.M. Haugan, J.A. Johannessen, Nansen Rem. Sens. Cen., Bergen, Norway, M. Ikeda, Bedford Inst. of Ocean., Dartmouth, NS, Canada, K. Kloster, Nansen Rem. Sens. Cen., Bergen, Norway | 736 |
| Analysis of Geosat altimeter sea state data and comparison with simultaneous aircraft SAR observations for NORCSEX-88 N.M. Mognard, Cen. Nat. d'Etudes Spat., Toulouse-Cedex, France, C. Livingstone, CCRS, Ottawa, Ont., Canada, J.A. Johannessen, P. Samuel, NRSC, Bergen, Norway, R. Shuchman, ERIM, Ann Arbor, MI, USA | 737 |
| Geosat derived winds, waves, and swells in the North Atlantic during NORCSEC-88 ; N.M. Mognard, Cen. Nat. d'Etudes Spat., Toulouse-Cedex, France, W.J. Campbell, E.G. Josberger, Univ. of Puget Sound, Tacoma, WA, USA | 742 (over) |
| A study of textural and tonal information for classifying sea ice from SAR imagery T. Hirose, Noetix Research Inc., Ottawa, Ont., Canada, S. Paterson, Intera Tech. Ltd., Ottawa, Ont., Canada, L. McNutt, CCRS, Ottawa, Ont., Canada | 747 |
| Ice classification algorithm development and verification for the Alaska SAR facility using aircraft imagery B. Holt, R. Kwok, E. Rignot, JPL/CIT, Pasadena, CA, USA | 751 |
| Speckle reduction and maximum likelihood classification of SAR images from sea-ice recorded during MIZEX 87 E. Holbaek-Hanssen, H. Tjelmeland, Norwegian Computing Centre, Oslo, Norway, O.M. Johannessen, T. Olaussen, R. Karpuz, Nansen Rem. Sens. Cen., Bergen, Norway | 755 |
| Multivariate analysis of texture statistics for SAR sea ice discrimination D.G. Barber, E. LeDrew, Univ. of Waterloo, Waterloo, Ont., Canada | 759 |
| Texture measures for sea-ice classification from radar images M.E. Shokr, Atmos. Env. Serv., North York, Ont., Canada | 763 |
| Comparison of sea ice parameters retrieved from passive microwave (SSM/I), Landsat MSS and AVHRR imagery A.J. Schweiger, Univ. of Colorado, Boulder, CO, USA, K. Steffen, Swiss Fed. Inst. of Tech., Zurich, Switzerland | 769 |
| Sensitivity of passive microwave sea ice concentration algorithms to the selection of locally and seasonally adjusted tie points. K. Steffen, Swiss Fed. Inst. of Tech., Zurich, Switzerland, A. Schweiger, Univ. of Colorado, Boulder, CO, USA | 773 |

| | Page |
|--|------|
| Utilization of local texture transforms for adaptive filtering of SAR sea ice imagery D.F. Salter, E.F. LeDrew, Univ. of Waterloo, Waterloo, Ont., Canada | 777 |
| Spring surface circulation patterns detected using remote sensing of drifting ice floes in Hudson Bay, Canada P. Larouche, Min. des Peches et des Oceans, Mont-Joli, Que., Canada | 780 |
| Detailed vegetation mapping from MEXIS-II digital data P.M. Treitz, P.J. Howarth, R.C. Suffling, P. Smith, Univ. of Waterloo, Waterloo, Ont., Canada | 783 |
| Evaluation of classification results obtained with combined multitemporal optical and microwave data A. Fiumara, N. Pierdicca, Telespazio, Rome, Italy | 787 |
| Cover class mapping in agricultural environments N.A. Campbell, J.F. Wallace, CSIRO, Wembley, W.A., Australia | 791 |
| The use of multirate Landsat Thematic Mapper imagery for agricultural land use mapping applied to electric power transmission route selection studies R.N. Pierce, C.L. Wagner, Ontario Hydro, Toronto, Ont., Canada | 795 |
| The use of multi-temporal TM tasseled cap features for land use mapping in European marginal areas - an operational approach J. Hill, J. Megier, Joint Res. Cen., Ispra, Italy | 798 |
| Land cover analysis of multitemporal thematic mapper data S. Ueno, Kyoto School of Comp. Sci., Kyoto, Japan, Y. Kawata, T. Kusaka, Kanazawa Inst. of Tech., Ishikawa, Japan | 802 |
| Microwave remote sensing of land surface with airborne active and passive sensors P. Coppo, Centre for Mic. Rem. Sensing, Florence, Italy, P. Ferrazzoli, Univ. Tor Vergata, Rome, Italy, G. Luzi, Centre for Mic. Rem. Sensing, Florence, Italy, P. Pampaloni, IROE/CNR, Florence, Italy, G. Schiavon, D. Solimini, Univ. Tor Vergata, Rome, Italy | 806 |
| Identification and assessment of anomalies on radar imagery of crops G.H. Holder, P.J. Howarth, Univ. of Waterloo, Waterloo, Ont., Canada, R. Mussakowski, Ont. Cen. for Rem. Sens., North York, Canada | 810 |
| Methode efficace et peu couteuse de cartographie des cultures par teledetection satellitaire P. Lafrance, J.-M. Dubois, Univ. de Sherbrooke, Sherbrooke, Que., Canada, M. Carignan, Serv. de conserv. des ress. agric., Que. Canada | 813 |
| Evaluation of agricultural management applications in the Rio Grande valley using SPOT imagery A.J. Richardson, Remote Sensing Research Unit, Weslaco, Texas, USA, M.M. Crawford, Univ. of Texas at Austin, Austin, Texas, USA, J.H. Everett, Rem. Sens. Res. Unit, Weslaco, Texas, USA, L. Dron, Univ. of Texas at Austin, Austin, Texas, USA, R. Summy, Rem. Sens. Res. Unit, Weslaco, Texas, USA | 816 |
| Cooperative systems for perceptual tasks in a remote sensing environment A.K. Mackworth, Univ. of British Columbia, Vancouver, BC, Canada | 819 |
| Expert systems and environmental change information extraction from remote sensing data; \rightarrow (10 p xi) \wedge D.G. Goodenough, CCRS, Ottawa, Ont., Canada, M.A. Robson, INTERA Tech. Ltd., Ottawa, Ont., Canada, K.B. Fung, CCRS, Ottawa, Ont., Canada | 823 |

cont
1

| | Page |
|---|------|
| Multisource data analysis in remote sensing and Geographic Information Systems based on Shafer's theory of evidence H. Kim, P.H. Swain, Purdue Univ., W.Lafayette, IN, USA | 829 |
| Expert system design for radar reconnaissance of sea ice G.M. Shirliffe, Univ. of Kansas, Lawrence, KS, USA | 833 |
| Classification of multitemporal remotely sensed images based on a fuzzy logic technique P. Blonda, Inst. Elab. Segnali Immagini, Bari, Italy, R.L. Polosa, Tecno.-CSATA-Novus Ortus, Valenzano, Italy, S. Losito, Cen. di Geodesia Spaziale, Matera, Italy, A. Mori, Tecno.-CSATA-Novus Ortus, Valenzano, Italy, G. Pasquariello, Inst. Elab. Segnali Immagini, Bari, Italy, F. Posa, Univ. Bari, Bari, Italy, D. Ragno, Inst. Biolog.-Selvicol., Viterbo, Italy | 834 |
| Integration of remote sensing and geophysical/geological data using Dempster-Shafer approach W.M. Moon, Univ. of Manitoba, Winnipeg, Man., Canada | 838 |
| An expert system for using digital terrain models D.G. Goodenough, CCRS, Ottawa, Ont., Canada, B. Baker, INTERA Tech. Ltd., Ottawa, Ont., Canada, G. Plunkett, CCRS, Ottawa, Ont., Canada, D. Schanzer, INTERA Tech. Ltd., Ottawa, Ont., Canada | 842 |
| A knowledge based system for the interpretation of SAR images of sea ice J.G. McAvoy, E.M. Krakowski, INTERA Tech. Ltd., Calgary, Alta., Canada | 844 |
| A fuzzy expert system for remote sensing image analysis F. Wang, Univ. of Waterloo, Waterloo, Ont., Canada | 848 |
| ESSII: Expert System for Satellite Image Interpretation C.V S. Sarma, Central Univ., Hyderabad, India, R. Krishnan, Nat. Rem. Sens. Agency, Hyderabad, India, B.E. Prasad, A.K. Pujari, Central Univ., Hyderabad, India | 852 |
| SHERLOCK supports the geocoding of SAR images M. Ploessnig, W.G. Kropatsch, D. Strobl, Inst. for Image Processing and Comp. Graphics, Graz Austria | 856 |
| Atmospheric correction of high resolution land surface images D.J. Diner, J.V. Martonchik, E.D. Danielsion, C.J. Bruegge, NASA/JPL, Pasadena, CA, USA | 860 |
| Surface reflectance retrieval using atmosphere correction algorithms P.M. Teillet, CCRS, Ottawa, Ont., Canada | 864 |
| Image based atmospheric correction of airborne imagery J.R. Freemantle, J.R. Miller, York Univ., North York, Ont., Canada, N.T. O'Neill, Univ. de Sherbrooke, Sherbrooke, Que., Canada | 868 |
| Analyse des aerosols a partir des images AVHRR-NOAA 9 et 10. Application a la cartographie des sulfates L. Chartier, N.T. O'Neill, A. Royer, Univ. de Sherbrooke, Sherbrooke, Que., Canada | 869 |
| Removal of the scan angle effect on NOAA-AVHRR data in visible and near-IR parts of the electromagnetic spectrum C. Yang, A. Vidal, Lab. Commun Teledetec., CEMAGREF ENGREF, Montpellier, France, D. Lepoutre, GEOSYS, Toulouse, France, D. Pozzobon, ORMVAG, Kenitra, Morocco | 873 |
| A measurement program for the validation of atmospheric radiative transfer models R.P. Gauthier, F.J. Ahern, P.M. Teillet, CCRS, Ottawa, Ont., Canada | 877 |

| | Page |
|---|------|
| Aerosol anomalies in Nimbus-7 Coastal Zone Color Scanner data obtained in Japan area H. Fukushima, Tokai Univ., Orido, Japan, R.C. Smith, Univ. of California, Santa Barbara, CA, USA, Y. Sugimori, M. Toratani, Tokai Univ., Orido, Japan, Y. Yasuda, Chiba Univ., Chiba, Japan | 881 |
| The distribution of clear-sky radiation over varying terrain R. Dubayah, J. Dozier, F. Davis, Univ. of California, Santa Barbara, CA, USA | 885 |
| Evaluation of atmospheric effect on remotely sensed data based on measurement of spectral radiance Y. Yasuoka, T. Miyazaki, Y. Iikura, Nat. Inst. for Env. Studies, Ibaraki, Japan | 889 |
| Integrated atmospheric correction of multispectral satellite imagery B.C. Robertson, B. Sharpe, MacDonald Dettwiler, Richmond, BC, Canada, P.M. Teillet, CCRS, Ottawa, Ont., Canada | 893 |
| Atmospheric correction of NIMBUS-7 CZCS images using multiple scattering data base S. Mukai, Kanazawa Inst. of Tech., Ishikawa, Japan | 897 |
| Using LOWTRAN 6 and DEM to derive path radiance for SPOT imageries over mountainous terrain J.-Y. Chen, A.J. Chen, H.T. Wang, Nat. Cen. Univ., Chung-Li, Rep. China | 900 |
| Optical archiving for scientific data B. Lowrey, NASA/Goddard Space Flight Center, Greenbelt, MD USA | 901 |
| Optical storage systems as the media of remotely sensed data exchange J. Iisaka, CCRS, Ottawa, Ont., Canada | 904 |
| Landsat electron beam recorder P.F. Grosso, J.P. Whitley, Image Graphics, Inc., Shelton, CT, USA | 908 |
| Data compression of spaceborne remotely sensed imagery from coherent and incoherent sensor systems S.A. Mango, K.W. Hoppel, Naval Res. Lab., Washington, DC, USA | 912 |
| Data compression scheme for archiving of ADEOS data K. Arai, NASDA, Saitama, Japan | 914 |
| Noiseless image compression K. Miettinen, N. Beser, Gen. Electric, King of Prussia, PA, USA | 919 |
| Data compression on the Mars Rover Sample Return mission N. Beser, K. Miettinen, Gen. Electric, King of Prussia, PA, USA | 921 |
| IHS transform for the integration of radar imagery with geophysical data J. Harris, R. Murray, INTERA Tech. Ltd./CCRS, Ottawa, Ont., Canada | 923 |
| Polarization diversity radars in meteorology J.I. Metcalf, Air Forces Geoph. Lab., MA, USA | 927 |
| The differential reflectivity dual-polarization radar technique: accomplishments and future prospects in meteorology T.A. Seliga, Pennsylvania State Univ., Univ. Park, PA, USA | 928 |
| Remote sensing of random media with ellipsoidal inhomogeneities A. Sihvola, I.V. Lindell, Helsinki Univ. of Tech., Espoo, Finland | 929 |

| | Page |
|--|------|
| Use of circular polarization in a marine radar positioning system D.C. Michelson, D.J. MacNeil, E.V. Jull, Univ. of British Columbia, Vancouver, BC, Canada, R.G. Lyall, H.H. Lanziner, Offshore Sys. Ltd., N.Vancouver, BC, Canada | 932 |
| The frequency and polarization dependence of complex RCS signatures S. Riegger, D. Kahny, W. Wiesbeck, Univ. Karlsruhe, Karlsruhe, FR Germany | 936 |
| Effect of target size and tilt on target identification based upon full polarimetric radar imaging data J.R. Huynen, PQ Research, Los Altos Hills, CA, USA | 937 |
| Optimization procedures for scattering matrices in the coherent and partially coherent cases W.-M. Boerner, W-L Yan, A.B. Kostinski, Univ. of Illinois, Chicago, IL, USA | 941 |
| Recent advances in airborne terrestrial remote sensing with the NASA AVIRIS, Airborne SAR, and TIMS G. Vane, D.L. Evans, A.B. Kahle, NASA/JPL, Pasadena, CA, USA | 942 |
| Mineral identification by the AVIRIS data I. Pippi, CNR - IROE, Firenze, Italy | 944 |
| A study of the Dolly Varden mountains, Nevada, through the use of broad band remote sensing and imaging spectrometry J.A. Zamudio, W.W. Atkinson Jr., Univ. of Colorado, Boulder, CO, USA | 948 |
| Mapping hydrothermally altered rocks with the Airborne Imaging Spectrometer (AIS) and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) \rightarrow (to p XIV) \rightarrow F.A. Kruse, D.L. Taranik, Univ. of Colorado, Boulder, CO, USA | 952 |
| Lithology and structure within the basement terrain adjacent to Clark Mountains California mapped with calibrated data from the Airborne Visible/Infrared Imaging Spectrometer R.O. Green, G. Vane, NASA/JPL, Pasadena, CA, USA | 957 |
| Comparison of Landsat Thematic Mapper and geophysical and environmental research imaging spectrometer data for the Cuprite mining district, Esmeralda and Nye counties, Nevada K.S. Kierein-Young, F.A. Kruse, Univ. of Colorado, Boulder, CO, USA | 961 |
| Identification of illite polytype zoning in disseminated gold deposits using reflectance spectroscopy and X-ray diffraction - potential for mapping with imaging spectrometers F.A. Kruse, P.L. Hauff, Univ. of Colorado, Boulder, CO, USA | 965 |
| Present status and operation history of MOS-1 K. Arai, NASDA, Saitama, Japan | 969 |
| Interband correlations of visible and thermal infrared radiometer imagery from the Marine Observation Satellite-1 M. Heric, Autometric, Inc., Alexandria, Virginia, USA | 974 |
| Verification results of MOS-1 Microwave Scanning Radiometer (MSR) K. Maeda, M. Notomo, S. Ogawa, H. Sato, Nat. Sp. Dev. Agency of Japan, Saitama-ken, Japan | 978 |
| The Canadian MOS-1 project T. Butlin, M. Manore, R.A. O'Neil, CCRS, Ottawa, Ont., Canada | 982 |

| | Page |
|--|------|
| Geometric and radiometric correction of MOS-1 imagery in a Canadian processing system A. Erickson, CCRS, Ottawa, Ont., Canada, B. Robertson, K. Wiebe, MacDonald Dettwiler, Richmond, BC, Canada | 987 |
| A study on sea ice monitoring using MOS-1/MSR K. Cho, K. Takeda, Rem. Sens. Tech. Cen. of Japan, Tokyo, Japan, K. Maeda, H. Wakabayashi, Earth Obs. Cen., Saitama-kan, Japan | 991 |
| An evaluation of MOS-1 MESSR data for land applications W.M. Strome, PCI Inc., Richmond Hill, Ont., Canada, E.F. LeDrew, Univ. of Waterloo, Waterloo, Ont., Canada | 995 |
| The design and implementation of the European Radar Cross Section database (EURACS) I.R. Barnes, Hunting Tech. Serv., Herts, UK, D. Brewster, N.J. Veck, GEC-Marconi Res. Cen., Essex, UK, A. Sieber, Ispra Est., Ispra, Italy | 996 |
| Digital simulation of terrain backscattering cross sections A.W. Biggs, Univ. of Alabama, Huntsville, AL, USA | 1000 |
| Object orientated design demonstrated by SAR raw data simulator development A. Popella, R. Konjack, F. Lesske, DFVLR, Oberpfaffenhofen, FR Germany | 1004 |
| Noise modeling and estimation of remotely-sensed images J.S. Lee, K. Hoppel, Naval Res. Lab., Washington, DC, USA | 1005 |
| Simulation of subpixel terrain effects on radar backscattering of snow J. Shi, J. Dozier, Univ. of California, Santa Barbara, CA, USA | 1009 |
| Calculation of the spatial distribution of scatterers in a diffuse scene from SAR data C.C. Wackerman, Env. Res. Inst. of Mich., Ann Arbor, MI, USA | 1012 |
| Combating speckle in SAR images: Vector filtering and sequential classification based on a multiplicative noise model Q. Lin, J. Allebach, Purdue Univ., W. Lafayette, IN, USA | 1016 |
| Visual effect of speckle reduction on interpretation of one-look SAR photo images H. Kimura, N. Motomura, N. Kodaira, Rem. Sens. Tech. Cen. of Japan, Tokyo, Japan, H. Hirowawa, Inst. Space & Astr. Science, Kanagawa, Japan | 1020 |
| The FLI airborne imaging spectrometer: experience with land and water targets J.F.R. Gower, Inst. of Ocean Sciences, Sidney, BC, Canada, R.A.H. Buxton, Moniteq Ltd., Concord, Ont., Canada, G.A. Borstad, G.A. Borstad Ass. Ltd., Sidney, BC, Canada | 1024 |
| A Compact Airborne Spectrographic Imager (CASI) S.K. Babey, C.D. Anger, ITRES Res. Ltd., Calgary, Alta., Canada | 1028 |
| Imaging spectrometry at the focal plane T.S. Pagano, C.F. Schueler, L.M. Woody, Santa Barbara Res. Cen., Goleta, CA, USA | 1032 |
| Quantitative determination of imaging spectrometer specifications based on spectral mixing models A.F.H. Goetz, J.W. Boardman, CIRES, Univ. of Colorado, Boulder, CO, USA | 1036 |
| Fluorescence lidar design: optimization criteria G. Cecchi, L. Pantani, M. Romoli, Consiglio Naz. delle Ricerche, Firenze, Italy | 1040 |
| Multi-field-of-view lidar for single ended measurement of extinction coefficients P.E. LaRocque, A. Ulitsky, Optech Inc., Downview, Ont., Canada | 1043 |

| | Page |
|---|------|
| Synchronous fluorescence spectroscopy of dissolved organic matter to optimize lidar detection parameters A. Vodacek, Cornell Univ., Ithaca, NY, USA | 1046 |
| Lidar studies of atmospheric extinction S.R. Pal, A.I. Carswell, A.G. Cunningham, York Univ., North York, Ont., Canada | 1050 |
| The NOAA Geosat program: Monitoring tropical sea level with satellite altimetry R. Cheney, B. Douglas, L. Miller, R. Agreen, N. Doyle, NOAA, Rockville, MD, USA | 1053 |
| Geoid estimates in the Gulf Stream from Geosat altimetry data and a Gulfcast mean sea surface D.L. Porter, Johns Hopkins Univ., Laurel, MD, USA, S. Glenn, A. Robinson, Harvard Univ., Cambridge, MA, USA | 1054 |
| Sea level variations of the N. E. Atlantic measured by Geosat J. Thomas, Univ. Coll. of N. Wales, Gwynedd, UK, P. Woodworth, Bidston Observ., Merseyside, UK | 1058 |
| Sea-surface height anomalies in the north-east Pacific as observed with the Geosat altimeter: the Sitka eddy J.F.R. Gower, Institute of Ocean Sciences, Sidney, BC Canada | 1059 |
| Mesoscale variability off California as seen by the Geosat altimeter P. Flament, Univ. of Hawaii, Honolulu, HI, USA, P.M. Kosro, A. Huyer, Oregon State Univ., Corvallis, OR, USA | 1063 |
| Eddy scales resolved by the Geosat radar altimeter J.J. Bisagni, US Naval Underwater Sys. Center, Newport, RI, USA | 1069 |
| Analysis of the fetch-related bias in altimeter wind speed measurements R.E. Glazman, S.H. Pilorz, NASA/JPL, Pasadena, CA, USA | 1074 |
| Global altimeter measurements of extremes of wind speeds and waveheights and their impact on forecasting E.B. Dobson, Johns Hopkins Univ., Laurel, MD, USA | 1077 |
| Satellite altimetry: Utilization for resource exploration D. Delikaraoglou, Cdn. Centre for Surveying, Ottawa, Ont., Canada, J.A.R. Blais, Univ. of Calgary, Calgary, Alta., Canada, E.R. Kanasevich, Univ. of Alberta, Edmonton, Alta., Canada, D. Agouridis, Geoscience Integ. Inc., Edmonton, Alta., Canada | 1078 |
| NORCSEX wind stress measurements from a ship and a buoy K.L. Davidson, C. Skupniewicz, Naval Postgrad. Sch., Monterey, CA, USA, O. Skagseth, Nansen Rem. Sens. Cen., Bergen, Norway | 1082 |
| Scatterometer measurements of wind, waves and ocean fronts during NORCSEX R.G. Onstott, R.A. Shuchman, Env. Res. Inst. of Mich, Ann Arbor, MI, USA, J.A. Johannessen, O. Skagseth, Nansen Rem. Sens. Cen., Bergen, Norway, K. Davidson, Naval Postgrad. Sch., Monterey, CA, USA | 1084 |
| Correlation of marine radar cross section statistics with wind stress during NORCSEX D.B. Trizna, Naval Res. Lab., Washington, DC, USA, K. Davidson, Naval Postgrad. Sch., Monterey, CA, USA | 1089 |
| Scalar winds from SSM/I in the Norwegian and Greenland Seas during NORCSEX P. Gloersen, NASA/GSFC, Greenbelt, MD, USA, P. Hubanks, Res. & Data Sys. Corp., Lanham, MD, USA | 1090 |

| | Page |
|---|------|
| Evaluation of Geosat altimeter wind and wave data L.-A. Breivik, The Norwegian Meteorological Institute, Oslo, Norway | 1094 |
| C-Band synthetic aperture Radar measurements of moving ocean waves C.L. Rufenach, A.O. Scheffler, R.A. Shuchman, Env. Res. Inst. of Mich, Ann Arbor, MI, USA, J.A. Johannessen, Nansen Rem. Sens. Cen., Bergen, Norway | 1098 |
| Wave measurements on Haltenbanken during NORCSEX '88: An intercomparison of buoy, SAR and altimeter data R.B. Olsen, S.F. Barstow, H. Schyberg, Occan. Co. of Norway, Trondheim, Norway | 1104 |
| Experimental investigation of imaging geometry effects on SAR-derived ocean wave spectra K.A. Hogda, J.P. Pedersen, T. Gunerlussen, H. Johnsen, FORUT, Tromsø, Norway, T. Eltoft, Univ. of Tromsø, Tromsø, Norway | 1108 |
| Automated tracking of Arctic ice floes in multitemporal SAR imagery R. McConnell, W. Kober, F. Leberl, Vexcel Corp., Boulder, CO, USA, R. Kwok, J. Curlander, NASA/JPL, Pasadena, CA, USA | 1112 |
| Automated analysis of polar satellite imagery J. Banfield, Montana State Univ., Bozeman, MT, USA, D. Rothrock, Univ. of Washington, Seattle, WA, USA | 1117 |
| Object-based feature-tracking algorithms for SAR images of the marginal ice zone J. Daida, J. Vesecky, Stanford Univ., Stanford, CA, USA | 1121 |
| A Hough transform technique for extracting lead features from sea ice imagery F.M. Fetterer, R.J. Holyer, Stennis Space Center, MS, USA | 1125 |
| Extraction of ridge feature characteristics from SAR images of sea ice J.F. Vesecky, M.P. Smith, R. Samadani, Stanford Univ., Stanford, CA, USA | 1129 |
| Automatic estimation of ice kinematics using remote sensing data H. Flesche, K. Kloster, T. Olaussen, O.M. Johannessen, Nansen Rem. Sens. Cen., Bergen, Norway | 1133 |
| Analysis of ice motion vectors determined by the ASF GPS for Seasat SAR observations in the Beaufort Sea F. Carsey, NASA/JPL, Pasadena, CA, USA | 1137 |
| Preliminary observations of Labrador Sea marginal ice zone rheology using C-Band SAR M.R. Drinkwater, NASA/JPL, Pasadena, CA, USA, V.A. Squire, Univ. of Otago, New Zealand | 1138 |
| Evaluation of ice conditions in the Northumberland Strait using C and X band SAR imagery, (EDC) J.A. Dechka, V.L. Shaw, F.G. Bercha, T.G. Brown, The Bercha Group, Calgary, Alta., Canada | 1142 |
| Measured ice floe sizes in the Beaufort Sea W.J. Stringer, Univ. of Alaska Fairbanks, Fairbanks, AK, USA | 1147 |
| Ice kinematic measurements from aerial photography T.G. Brown, J.A. Dechka, J.W. Steen, The Bercha Group, Calgary, Alta., Canada | 1148 |
| The determination of ice displacements from sequential SAR imagery K.D. Oliphant, J.F. Sykes, E.D. Soulis, Univ. of Waterloo, Waterloo, Ont., Canada | 1152 |

COMPARISON OF THEORETICAL AND MEASURED BRIGHTNESS TEMPERATURE OF A SOYBEAN CANOPY

R.H. Lang¹, D.M. Le Vine
P. O'Neill, S. Saatchi
NASA Goddard Space Flight Center
Greenbelt, MD 20771 USA

O. Yazici
George Washington University
Washington, D.C. 20052 USA

T. Jackson
US Department of Agriculture
Beltsville, MD 20705 USA

ABSTRACT

Research is being performed to develop a model for the brightness temperature of a canopy which is based on the physical characteristics of the vegetation. This is done by employing the Peake model to relate the scattering coefficient of the vegetation to its radiometric brightness temperature and then by using a discrete scattering model to compute the scattering coefficient. A distorted Born approximation is used to provide a tractable solution to the scattering problem.

In order to verify the theory, a comparison has been made with experiments. The measurements were performed at the USDA Agricultural Research Center in Beltsville, Maryland in early September, 1987. Radiometric measurements of brightness temperature were made at L- and C-band over a crop of mature soybeans along with ground truth measurements. All model input parameters were measured; there were no "free" parameters to adjust in order to obtain agreement with the measurements.

Results showed that agreement between theory and measurements was reasonable for horizontal polarization at both L- and C-bands with dry underlying soil. Vertically polarized measurements also matched theory at C-band; however, the L-band theory did not track the data closely.

Key words: brightness temperature, microwave, vegetation

INTRODUCTION

In this paper microwave thermal emission from a soybean canopy is modeled by a layer of discrete scatterers over a flat lossy ground. Leaves in the canopy are represented by thin lossy dielectric discs which have random orientation statistics while stems are neglected. The brightness temperature is computed by using the relationship between passive and active problems [Peake, 1951, Tsang et al, 1982].

The active problem considered is that of a unit magnitude plane wave incident on the canopy. The method makes use of the fact that the plane wave energy absorbed by the canopy is equal to the emission from the canopy at the angle of plane wave incidence under conditions of thermodynamic equilibrium. Using energy conservation the absorbed energy is directly related to the energy scattered from the canopy surface. Thus the passive problem is reduced to the

¹ - On sabbatical leave from The George Washington Univ.

active problem of computing all the scattered energy from the canopy surface.

The plane wave energy scattered from the canopy surface consists of a specularly reflected portion from the underlying ground which is attenuated by the canopy and a portion arising from scattering by the

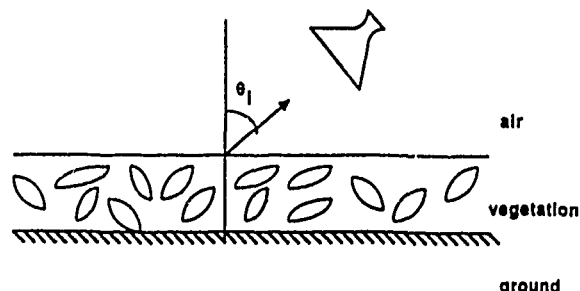


Fig. 1. Soybean canopy model

leaves. The mean or average wave solution in the canopy region is used to compute the specular portion while the hemispherical surface integral over the bistatic scattering coefficients gives the contribution due to scattering from the leaves. Here the distorted Born approximation is used to compute the bistatic scattering coefficients.

The topic of emission from vegetation canopies has been treated by many previous researchers in the context of remote sensing of soil moisture in the presence of a vegetation cover. Basharinov and Shutko, [1975] and Kirdiashev et al, [1979] treated the vegetation as an absorptive layer using the assumption that the scatterers have a low albedo. This model has been used by Jackson et al, [1982] to compare measured and model results. Their model parameters were roughly estimated using ground truth data. This same model has been used by Ulaby et al, [1984] to match data from a vegetation covered field with screens placed between the rows to isolate surface and vegetation scattering effects. Mo et al, [1982] modeled the canopy by solving the scalar transport equations under the assumption that the scatterers had peaked scattering patterns in the forward scattering direction. This theory had reasonable agreement with experiment but has a number of free parameters that can be adjusted. Other transport theory solutions of the continuous or discrete type have been made by Wang

et al [1984], Eom and Fung, [1985] and Tsang et al, [1985]. However, in all cases the ground truth information has never completely specified all the model parameters thus leaving certain parameters free to adjust within certain limits. Other references are contained in a review by Schmugge et al, [1986].

In the following section the details of the passive model will be specified and related to the corresponding active problem. The solution to the active problem will be given, thus completing the model description. In the next section the experiment will be discussed including a summary of the ground truth data collection and analysis. In the final section, a comparison between the calculated results and the measured data will be made.

Model Formulation

A mature soybean canopy of thickness d is modeled by a layer of lossy dielectric discs having random position and orientation statistics. It is assumed that the discs are uniformly distributed in the azimuthal coordinate and that they are randomly placed throughout the layer. The ground under the vegetation is represented by a lossy dielectric half space (Fig. 1). Thus rough surface effects and effects due to an open canopy structure are not treated by this model.

The emission from the vegetation and underlying ground is observed by a radiometer whose look direction makes an angle θ_1 with the normal to the layer as is shown in Fig. 1. It is assumed that the canopy, air and ground are all at temperature T and thus are in thermodynamic equilibrium. The brightness temperature T_{Bq} , where q is the polarization of the received energy, can be normalized to the physical temperature of the layer by introducing the emissivity, e_q , as follows:

$$T_{Bq} = e_q T, \quad q = h, v \quad (1)$$

Here h and v stand for horizontal and vertical polarization. As defined here the emissivity is associated with the upper surface of the canopy.

Although the emissivity can be computed using transport theory, it is instructive to employ the Peake, [1951] method relating active and passive problems. Consider the analogous active problem having a plane wave of polarization q incident upon a vegetation layer at an angle θ_1 with respect to the layer normal. If the plane wave has unit power density, then that portion of the power over a unit area absorbed by the layer and underlying ground is called the absorption coefficient, $a_q(\theta_1)$. For bodies in thermodynamic equilibrium, the power absorbed by a surface must be equal to the power emitted from it. It follows that

$$a_q(\theta_1) = e_q(\theta_1) \quad (2)$$

If the absorption coefficient for the active problem can be determined, then the emissivity for the passive problem will be known. The calculation of the absorption coefficient can be simplified by relating it to the scattering albedo, W_q , of the upper canopy surface. By using conservation of energy, one can readily show that

$$a_q(\theta_1) = 1 - W_q(\theta_1) \quad (3)$$

and thus the scattering albedo and absorption

coefficient are directly related.

To calculate the scattering albedo, the amount of power scattered into the upper half space must be computed when a plane wave of unit power density is incident upon the vegetation layer. This scattered power consists of two parts: the specular and the diffuse components. The specular component arises from

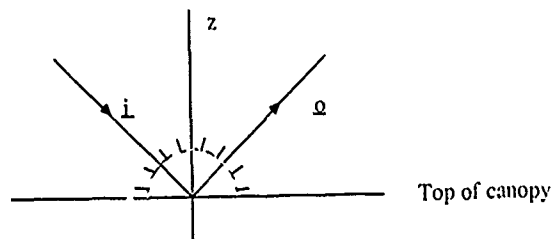


Fig. 2. Canopy specular and diffuse radiation

the incident wave being reflected from the underlying ground. It is attenuated as it passes through the vegetation due to the lossy nature of dielectric discs. The diffuse component results from waves that are scattered by the discs. This scattering results in waves exiting the layer into the air region in all possible directions.

Quantitatively, the scattering albedo is broken up into two components:

$$W_q(\theta_1) = W_{spq}(\theta_1) + W_{dfq}(\theta_1) \quad (4)$$

where W_{spq} is the specular component and W_{dfq} is the diffuse component. The specular component of the albedo is equal to the reflectivity, Γ_{sq} , of the vegetation layer. By using the expression for the mean wave [see Lang and Saleh, 1985] the surface reflectivity can be directly related to the reflectivity, Γ_{gq} , of the ground and the attenuation of the vegetation. The expression is:

$$W_{spq} = \Gamma_{sq} = \Gamma_{gq} e^{-2\alpha_q d} \quad (5)$$

where α_q is the extinction coefficient for polarization q ,

$$\alpha_q = \rho \sigma_{tq} \sec \theta_1, \quad (6)$$

ρ is the density of scatterers and σ_{tq} is total scattering cross section of a disc averaged over orientation. The factor of 2 in the exponent of (5) indicates that the wave is attenuated both going into the layer and coming out of the layer.

The diffuse component of the albedo can be obtained [see Peake, 1951] by integrating over the bistatic scattering coefficients as follows:

$$W_{dfq} = \frac{1}{4\pi \cos \theta_1} \int_{2\pi} d\Omega (\sigma_{hq}^o + \sigma_{vq}^o) \cos \theta_s \quad (7)$$

Here $\sigma_{pq}^o(o, i)$ is the bistatic scattering coefficient for the vegetation layer where the incident wave has direction i and polarization q while the scattered wave has direction o and polarization p . From (7), it is seen that both like and cross polarized components are integrated over an upward facing hemisphere (Fig. 2).

The distorted Born approximation [see Lang, 1981] can be used to obtain approximate expressions for the bistatic scattering coefficients which should be valid for L- and C-band frequencies. The distorted Born results divide naturally into three physically distinct contributions: a direct or volume scatter term, σ_{pqd}^o , a direct-reflected or interference term, σ_{pqdr}^o , and finally a reflected term, σ_{pqr}^o . In equation format this becomes:

$$\sigma_{pq}^o = \sigma_{pqd}^o + \sigma_{pqdr}^o + \sigma_{pqr}^o \quad (8)$$

The last term is usually negligible unless the reflectivity of the underlying surface is close to 1 which is not generally the case for soil. As a result, this term will be neglected. Following a procedure similar to that found in Lang and Sidhu [1983], expressions for the direct and direct-reflected components can be found by integrating over all single scattering contributions of particles embedded in an equivalent disc medium. The direct contribution is given by:

$$\sigma_{pqd}^o = \rho \sigma_{pqd} \frac{1 - e^{-(\alpha_p + \alpha_q)d}}{\alpha_p + \alpha_q} \quad (9)$$

where α_q is the attenuation coefficient for incident polarization q as given by (6) and α_p is the attenuation coefficient for the scattered wave of polarization p. Here (6) again can be used with q replaced by p and θ_i replaced by θ_s .

The direct-reflected contribution is found to be:

$$\sigma_{pqdr}^o = \rho \sigma_{pqdr}^+ \Gamma_{sq} \frac{1 - e^{-(\alpha_p - \alpha_q)d}}{\alpha_p - \alpha_q} + \rho \sigma_{pqdr}^- \Gamma_{sp} \frac{e^{-(\alpha_p - \alpha_q)d} - 1}{\alpha_p - \alpha_q} \quad (10)$$

where Γ_{sq} and Γ_{sp} are the canopy surface reflectivity for incident and scattered directions. It is given explicitly for the incident direction by (5). In equations (9) and (10) the scattering cross sections σ_{pqd}^o , σ_{pqdr}^+ and σ_{pqdr}^- are given in terms of 4π times the average absolute value of the scattering amplitude squared at the appropriate incident and scattering angles.

As mentioned previously, the leaves in the canopy have been modeled by thin dielectric discs having average radius a and thickness T. These parameters are measured in the field along with the leaf dielectric constant ϵ_r . By employing a quasi-static approximation which is valid when the leaf is thin compared to a wavelength, simple expressions for the bistatic scattering amplitude of the leaf can be obtained. The attenuation coefficients and scattering cross sections appearing in (5), (9) and (10) can be computed by taking the appropriate orientation average over the scattering amplitude. An expression for the scattering amplitude for a disc is given in Le Vine et al [1985].

MEASUREMENTS

A series of measurements has been conducted in early September 1987 at a test site at the USDA

Beltsville Agricultural Research Center. The data collected have been used to see how the measurements compare with the results obtained from the model discussed in the previous section. The setup includes L-band (1.4 GHz, 21 cm) and C-band (5.0 GHz, 6 cm) truck mounted radiometer systems. Brightness temperatures for both horizontal and vertical polarizations for incident angles from 10° to 60° in 10° steps have been measured. The nominal height of the radiometer from the canopy surface is 6 m which provides a footprint of almost 1.5 m in diameter.

Past attempts to compare measurements with theory have ignored the structural properties of the canopy, and free parameters have been used in the models in order to fit the measured data. In this paper, however, ground truth data has been used to set all the parameters in the model.

The data has been taken over a canopy of mature soybeans. The soybeans are planted in a uniform fashion to avoid the row structure effects. This has resulted in soybeans with leaves which are almost uniformly distributed in the azimuth angle. The data includes orientation and size distributions of leaves, volume density of leaves, canopy height and the dielectric constant of leaves and the underlying soil. Several small regions in the field have been chosen for collecting ground truth data in order to have a representative sample. The orientation and size distributions are depicted in Figures 3 and 4.

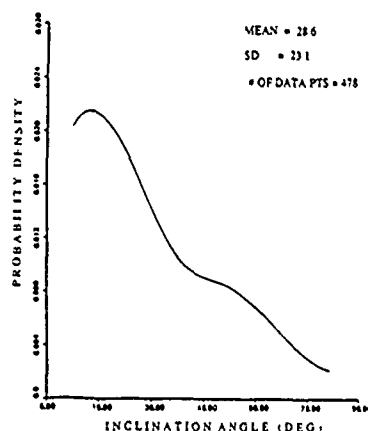


Fig. 3 Soybean leaf inclination angle distribution

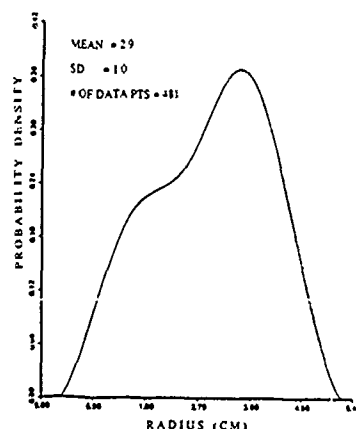


Fig. 4. Soybean leaf radius distribution

The orientation of the leaf is defined by inclination and azimuth angles. The mean inclination angle is 29° degrees which is the angle between the normal to the leaf surface and normal to the ground surface. This means that leaves are mostly parallel to the ground and slightly pointing upward. The azimuth angle which is measured with respect to a reference direction in the canopy is considered uniform. In calculating the radius distribution, the area of a leaf has been used to obtain the radius of an equivalent circular disc. Other measured quantities used in the model are as follows:

- $a = 2.9$ cm - average disc radius
- $T = 0.2$ mm - disc thickness
- $d = 0.8$ m - height of canopy
- $\rho = 2050$ / m³ - density of leaves

The plants, in their mature state, cover the canopy without leaving large areas of open ground. The average density of plants is 11.0 plants/m², and the leaf area index (LAI) has been calculated to be 4.85.

The dielectric constant of the underlying soil has been measured by a portable dielectric probe at L-band. The same probe has been used to obtain the dielectric constant of a stack of leaves. The instrument has the capability of giving average values of many sample readings. These values have been used in conjunction with the curves in Hallikainen et al, [1985] and Ulaby et al, [1987] to obtain the volumetric moisture and consequently the dielectric constants at C-band. It has been found that the underlying soil has a volumetric moisture of about 0.05 cm³/cm and for such a dry soil the changes in the real and imaginary parts of the dielectric constant are negligible. In the leaf case however, the dielectric constant changes from 23 ± 16 at L-band to 18 ± 15 at C-band.

RESULTS

The horizontal and vertical brightness temperatures versus incident angle are depicted in Figs. 5 and 6. In each figure the experimental and theoretical values are compared. The theoretical results are obtained by directly inserting the ground truth data in the model. The figures show that for both L-band and C-band the theoretical and experimental values are in reasonable agreement. It is observed that there is a maximum difference of approximately $3 K^\circ$ between measurement and theory in the C-band horizontal case and they both have the same trend. In the vertical case however, the difference becomes larger, up to $7 K^\circ$, at higher angles. This is due to the Brewster angle effect in the theoretical result which appears in a slightly upward trend whereas, the measurement stays flat (Fig. 6). This effect which is more noticeable at L-band (Fig. 5), up to $15 K^\circ$, might be related to the status of the radiometer system and the measurement techniques. These effects are still under investigation.

In Figs. 7 and 8, the theoretical result for horizontally polarized brightness temperature has been decomposed into the various scattering contributions. There are five curves in each figure and they correspond to TB_{exp} , experimental values, TB_0 , half space result, TB_{sp} , half space result attenuated by the layer, TB_{sd} , direct scattering term included, and TB_{sdr} , direct-reflected term included. It is seen that at L-band the effects of the scattering terms are minimal and the spectral return which includes the reflection from the soil and the attenuation through

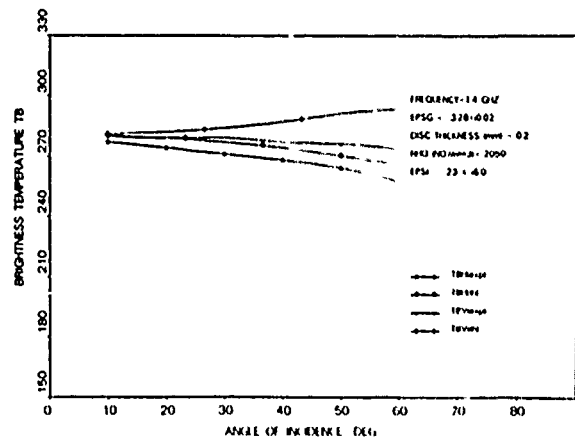


Fig. 5. Measured and calculated brightness temperatures at 1.4 GHz

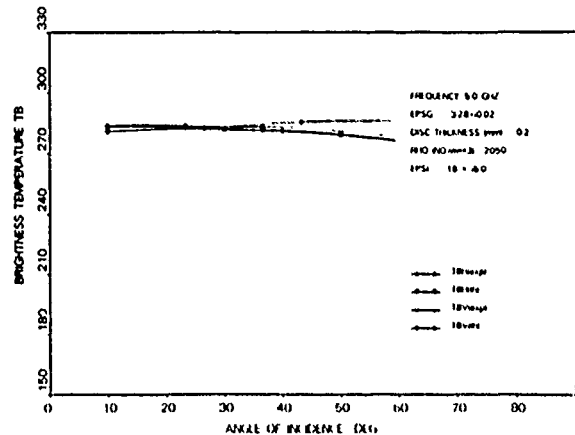


Fig. 6. Measured and calculated brightness temperatures at 5 GHz

the canopy dominates. At C-band (Fig. 8), the direct scattering becomes important and has a contribution of almost $10 K^\circ$. The direct reflected term, however, is not important at either frequency because of the low reflectivity of the ground. This result indicates that at L-band the scattering from the vegetation is negligible and the brightness temperature measured by the radiometer is mainly the emission from the ground. As the frequency increases the effect of the scattering from the layer becomes important.

It is important to mention that our result shows the same radiometric behavior of vegetation layers as predicted by the empirical model suggested by Kirdiashev et al, [1979]. In their model the emissivity of the soil is multiplied by the attenuation function of the canopy. In the model described in this paper if one includes only the specular return in equation (4) and substitutes (5) into (3) and then (2), the result reduces to the following form at L-band:

$$e_q = 1 - e_{gq} e^{-2\tau_q} \quad q = h, v \quad (11)$$

where $\tau_q = \alpha d$ and $e_{gq} = 1 - \Gamma_{gq}$ are the optical

depth and the emissivity of the ground respectively. Note that the attenuation function, which has been found empirically in Kirdiashev et al, [1979], can be related to the various structural properties of the canopy in our model. This analysis shows that as the frequency increases and the effect of the scattering becomes important, the absorptive model becomes less valid.

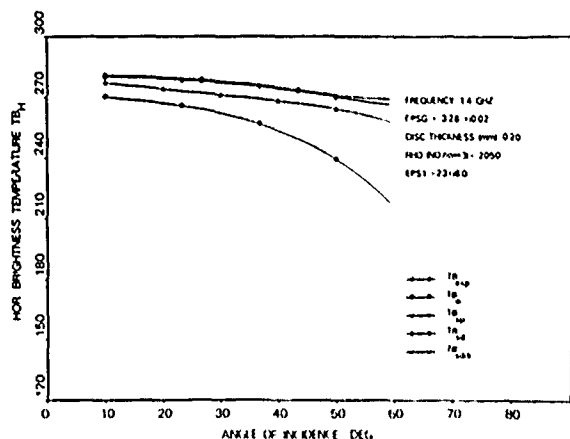


Fig. 7 Contribution of various scattering effects to L-band brightness temperature

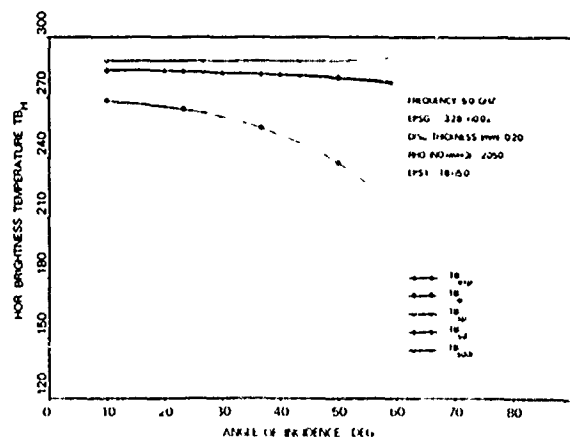


Fig. 8 Contribution of various scattering effects to C-band brightness temperature

REFERENCES

1. Basharinov A E & Shutko A M, "Simulation Studies of the SHF Radiation Characteristics of Soils and Under Moist Conditions," NASA Tech Transl. TT F-16, 1975.
2. Brunfeldt D R & Ulaby F T, "Measured Microwave Emission and Scattering in Vegetation Canopies," IEEE GE-22, S20-S24, 1984.
3. Eom H J & Fung A K, "A Scatter Model for Vegetation up to Ku-band," Remote Sensing of Environment 15, 185-200, 1985.
4. Hallikainen M T, Ulaby F T, Dobson M C, El-Rayes M A & Wu L, "Microwave Dielectric Behavior of Wet Soil - Part I," IEEE GE-23, 25-34, 1985.
5. Jackson T J, Schmugge T J & Wang J R, "Passive Microwave Sensing of Soil Moisture Under Vegetation Canopies," Water Resour. Res. 18, 1137-1142, 1982.
6. Kirdiashev K P, Chukhlantsev A A & Shutko A M, "Microwave Radiation of the Earth's Surface in the presence of vegetation cover," Radio Eng. Electron. 24, 256-264, 1979.
7. Lang R H, "Electromagnetic Scattering from a Sparse Distribution of Lossy Dielectric Scatterers," Radio Science 16, 15-30, 1981.
8. Lang R H & Sidhu J S, "Electromagnetic Scattering from a Layer of Vegetation: A Discrete Approach," IEEE GE-21, 62-71, 1983.
9. Lar H & Saleh H, "A Microwave Inversion of Leaf Area and Inclination Angle Distributions from Backscattered Data," IEEE GE-23, 685-694, 1985.
10. Le Vine D M, Schneider A, Lang R H & Carter H G, "Scattering from Thin Dielectric Disks," IEEE AP-33, 1410-1413, 1985.
11. Mo T, Choudhury B J, Schmugge T J, Wang J R & Jackson T J, "A Model for Microwave Emission from Vegetation-Covered Fields," J. Geophys. Res. 87(C13), 11229-11237, 1982.
12. Peake W H, "Interaction of Electromagnetic Waves with Some Natural Surfaces," IRE Trans. Antenna Prog. 7, S324-S329, 1951.
13. Schmugge T J, O'Neill P E & Wang J R, "Passive Microwave Soil Moisture Research," IEEE GE-24, 12-22, 1986.
14. Tsang L, Blanchard A, Newtow R & Kong J A, "A Simple Relation Between Active and Passive Microwave Remote Sensing Measurements of Earth Terrain," IEEE-GRS 20, 482-485, 1982.
15. Tsang L, Kong J A and Shin R T, Theory of Microwave Remote Sensing, John Wiley & Sons, New York, 1985.
16. Ulaby F T & El-Rayes M A, "Microwave Dielectric Spectrum of Vegetation - Part II: Dual Dispersion Model," IEEE GE-25, 550-557, 1987.
17. Wang J R, Shiue J C, Chuang S L, Shin R T & Dombrowski M, "Thermal Microwave Emission from Vegetated Fields: A Comparison Between Theory and Experiment," IEEE GE-22, 143-150, 1984.

Ku-Band Polarization Characteristics of Wheat and Canola

M.R. Hinds¹, A.G. Wacker¹, G.J. Sofko², M.P. Poettcker¹

¹Electrical Engineering Department

²Institute of Space and Atmospheric Studies, Physics Department
University of Saskatchewan, Saskatoon, Saskatchewan, S7N 0W0

ABSTRACT

A FM-CW Ku-band ground based scatterometer has been modified into a full polarimeter for experimental measurement of crop and soil polarization characteristics. This paper describes the polarimeter data collected from the test plot of wheat and one of canola during the 1987 Ground Microwave Operations Experiment in Saskatoon. The processing of this data into polarization state measurements is discussed and a graphical presentation is made of the polarization ellipse parameters (the orientation angle (ψ) and the ellipticity angle (χ)), and the polarization ratio (m). Significant variations in the polarization ratio as a function incidence angle, day number, and transmit polarization indicate that the plant material unpolarizes the backscattered Ku-Band radiation. There are no consistent variations in either of the ellipse angles, which suggests that there is no consistent depolarization of the backscattered radiation. The canola exhibits similar unpolarization signatures for both transmit polarizations, while the wheat exhibits a significant unpolarization signature for only the vertical transmit polarization. From these results, inferences are drawn about the scattering processes for wheat, and canola.

1. Introduction

Most recently, microwave remote sensing studies have focused on measuring the complete polarization state of the backscattered radiation from an active source. For example Van Zyl, Zebker, Evans et al. [1, 2] have processed SAR data collected using the JPL/NASA airborne SAR, into polarimetric scattering measurements to determine the backscatter properties of different terrain types and the variation in backscatter properties as a function of the transmitted elliptically polarized state. Based upon previous VHF radar polarimetry [3] at the University of Saskatchewan (U of S), the Ground Microwave Operations group (GMO) have converted the Ku-band transceiver of the Canadian Centre for Remote Sensing (CCRS) multi-band ground based scatterometer into a full polarimeter [4] in order to study the polarization characteristics of crops and soils.

2. The Ku-band Polarimeter

The CCRS ground-based transceivers are frequency modulated (FM), continuous wave (CW) radars, which transmit either horizontal (H) or Vertical (V) linear polarization states and receive both H and V linear states. The received H and V RF signals are mixed with the transmitted RF signal to produce sum and difference frequencies. The latter IF signal at audio frequencies, is sent to a 5% 22 kHz bandpass filter to pass only the power collected from approximately $\pm 2.5\%$ of the selected range (the range cell size). The actual range is set by adjusting the slope of the triangle wave used for FM modulation, which will create a difference frequency of 22 kHz for targets at that range. This basic FM radar hardware is illustrated in outlined (left half) of Fig. 1.

This basic scatterometer is converted into a polarimeter using the intensity method as outlined by Sofko [3] and later Poettcker [4]. The IF H and V signals, prior to power detection are the input to audio frequency summing networks, which generate outputs proportional to other selected intensities, as measured by different antennas. As derived by Born and Wolf [5], the Stokes parameters (s_0 , s_1 , s_2 , and s_3) can be expressed as the sum or difference of selected elliptical intensities (H - horizontal linear, V - vertical linear, X - linear at 45 degrees to H, Y - linear at 135 degrees to H, R - right circular, and L - left circular), where X and Y are the sum and difference of H and V IF signals, and R and L are the sum and difference of the H and the V (with a $\pi/2$ delay) IF signals. The original H and V and the four new intensity signals are input to linear power detectors which produce output voltages proportional to the root of the input power. Note that the summing process results in intensity measurements which are a factor of $\sqrt{2}$ larger than the original H and V channels, as reflected in equations 1 to 4 where the Stokes parameters are calculated from the intensities. The right side of Fig. 1 illustrates the added summing networks and the resulting four new intensity measurements.

$$s_0 = H^2 + V^2 \quad (1) \quad s_2 = \frac{1}{2} X^2 - \frac{1}{2} Y^2 \quad (3)$$

$$s_1 = H^2 - V^2 \quad (2) \quad s_3 = \frac{1}{2} R^2 - \frac{1}{2} L^2 \quad (4)$$

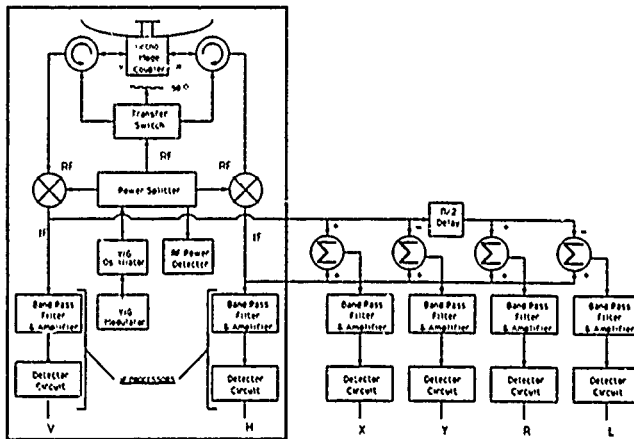


Figure 1: Polarimeter Functional Block Diagram

3. The 1987 GMO Experiment

In the 1987 GMO Experiment data was gathered from 34 one acre test fields of wheat, durham wheat, barley, canola and fallow. For this paper test plots #13 (hard red spring wheat) and #73 (canola) are selected for analysis. The experimental program was designed to observe each test plot at least twice each week through the growing season, where each observation produces a set of polarimeter measurements at ten incidence angles (10, 15, 20, 25, 30, 35, 40, 50, 60, and 70 degrees), at three row aspect angles (0, 45, and 90 degrees), at two transmitter polarizations (H,V) and at three range cells (5% of range centered at 95% of the desired range [cell 1], at 100% of the desired range [cell 2], and at 105% of the desired range [cell 3]). The raw polarimeter measurements consist of 1, 2, or 3 sets of ten trials (depending upon the number of independent samples per measurement as determined from crop height and incidence angle) of the 6 measured linear power detector output voltages, which are proportional to the root of the received power. The six measurements consist of three orthogonal intensity pairs: 1) H and V designate horizontal and vertical linearly polarized intensities, 2) X and Y designate linearly polarized intensities oriented at 45° and 135° with respect to horizontal, and 3) R and L designate the right and left circularly polarized intensities.

4. Data Processing and Calculation

The means and standard deviations of the measured voltages squared (each trial squared, then statistics calculated) represent the average intensity and the statistical error for each of the six intensity channels. As an example the mean intensity of H, \hat{I}_H , is calculated as

$$\hat{I}_H = \frac{1}{N_s} \sum_{i=1}^{N_s} (H_i)^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} I_{H_i} \quad (5)$$

$$\hat{I}_{pol} = \sqrt{(\hat{I}_H - \hat{I}_V)^2 + (\hat{I}_X - \hat{I}_Y)^2 + (\hat{I}_R - \hat{I}_L)^2} \quad (6)$$

$$\hat{m} = \frac{\hat{I}_{pol}}{(\hat{I}_H + \hat{I}_V)} \quad (7)$$

$$\hat{\psi} = \frac{1}{2} \text{ARCTAN} \left(\frac{\hat{I}_X - \hat{I}_Y}{\hat{I}_H - \hat{I}_V} \right) \quad (8)$$

$$\hat{\chi} = \frac{1}{2} \text{ARCSIN} \left(\frac{\hat{I}_R - \hat{I}_L}{\hat{I}_{pol}} \right) \quad (9)$$

These estimated intensities are used to calculate the average polarization state parameters:

- the average Stokes parameters as outlined in equations 1 through 4;
- or the average polarization ratio \hat{m} , the average orientation angle ($\hat{\psi}$), and the average ellipticity angle ($\hat{\chi}$).

The estimated average ellipse angles and polarization ratio are calculated as in equations 6 through 9.

5. The Polarization Characteristics of Wheat and Canola

For analysis, the received polarization state is compared with the transmitted state. Any changes in the polarization state, as shown by changes in m , ψ , and χ , are attributed to the target media. Two general categories of polarization effects are:

1. depolarization: the transfer of power from the elliptically polarized transmit state to another elliptical state, as identified by a change in ψ and χ ;
2. unpolarization: the transfer of power from the elliptically polarized transmit state to the unpolarized state, as denoted by a decrease in m or an increase in the unpolarized intensity, I_U .

The calculated polarization parameters are displayed in two formats:

- A) the parameter with error bars as a function of incidence angle (Fig. 2 for wheat and Fig. 3 for canola);
- B) the parameter on the Z axis of a 3 dimensional plot with time (in day number through the growing season) on the X axis and with incidence angle on the Y axis (Fig. 4 for wheat and Fig. 5 for canola).

The 2-D plots of m establish that significant decreases of the polarization ratio (or increases in the unpolarized

power component) occur in all crops as a function of incidence angle, transmit polarization and time. The 2-D plots of ψ and χ determine that both the orientation angle and ellipticity angles do not vary significantly from the transmitted polarization angles, which indicates that there is no consistent depolarization occurring due to the crops. The 3-D plots display more data on one plot for qualitative evaluation of the time/incidence angle polarization signatures for the different crops. Based on a review of Fig. 4 and Fig. 5 (for wheat and canola respectively), the most important observation is that significant unpolarization signature occurs only at vertical transmit polarization for wheat, while a different (and significant) unpolarization signature occurs at both transmit polarizations for canola.

From these observations, some general conclusions about the scattering properties of crop canopies are.

- crop canopies unpolarize the incident Ku-band radiation, as a function of time or development of the crop;
- crop canopies do not show any consistent depolarization characteristics;
- crop canopy geometry affects the sensitivity of the scattering process with respect to transmit polarization state orientation.

Previous studies by Le Toan et. al. [6] suggest that the loss factor for a medium is correlated with the volumetric water content of the canopy, and further that the loss factor for wheat is much larger for vertical transmit polarization than for horizontal transmit polarization. The loss factor for wheat also shows the same trend as the polarization ratio, or the unpolarized power content for wheat, indicating that the two parameters may be correlated. Ulaby [7] also indicates that the loss factor (attenuation) for soybeans measured and theoretical (calculated from the volumetric water content of the stems and heads) are well correlated. The polarization ratio signatures for wheat and canola show similar trends to those given by Ulaby (page 2114, Vol III, [7]) for soybeans as a function of time or crop development; as the crop grows the canopy moisture content increases and the polarization ratio decreases, but when the crop starts to mature and dry out, the polarization ratio increases back to approximately 1.0.

Another important conclusion drawn from these results is that because the primary polarization characteristic of canopies is unpolarization, standard cross polarization measurements σ_{HV}^0 or σ_{VH}^0 are primarily due to backscattered unpolarized power. These measurements are proportional to $(1/2)I_U$, the unpolarized power content of the channel, since there is only a small and inconsistent depolarization of co-polar power into cross polar power. Using the same line of reasoning the co-polar measurements σ_{HH}^0 or σ_{VV}^0 consist of both a polarized and an unpolarized component, $I_V = I_{pol} + I_U$. Given the transmit polarization sensitivity results for wheat established by Le Toan et. al., an explanation for the unpolarization effect observed at the vertical but not

at the horizontal transmit state is that the vertical stems of the wheat absorb a much larger portion of the polarized component at V than at H. The polarization ratio will decrease as a function of absorption. If the absolute magnitude of the unpolarized components at both transmit V and H are the same then the polarization ratio at V will be lower than at H.

6. Conclusions

In summary, it is clear that the measurement of the polarization characteristics of wheat and canola crop canopies is useful in developing a better understanding of the scattering properties of canopies in general. The crop canopies tend to unpolarize incident Ku-band elliptically polarized radiation with distinct time - incidence angle signatures. Furthermore, the unpolarization process is dependent upon the geometry of the crops, where a vertically oriented crop, such as wheat, exhibits much larger unpolarization at transmit V than at H and where a more randomly oriented crop, such as canola, exhibits similar unpolarization characteristics for both transmit polarizations. These results also suggest that the polarization ratio is negatively correlated (or the unpolarized power is positively correlated) with the canopy loss factor and volumetric water content.

References

1. Van Zyl, Jakob J., Zebker, Howard A., Elachi, Charles, "Imaging Radar Polarization Signatures: Theory and Observation", *Radio Science*, Vol. 22, No. 4, , July-Aug 1987, pp. 529-543.
2. Evans, Diane L., Farr, Tom G., Van Zyl, Jakob J., Zebker, Howard A., "Radar Polarimetry: Analysis Tools and Applications", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. GE-26, No. 6, November 1988, pp. 774-789.
3. Sofko, G. J., Kavadas, A., "Polarization of Auroral Backscatter at 42 MHz.", *Journal of Geophysical Research*, Vol. 76, 1971, pp. 1778-92.
4. Poettcker, M. P., "A K-Band Microwave Polarimeter", Master's thesis, University of Saskatchewan, 1987.
5. Born, M., and Wolf, E., *Principles of Optics*, Pergamon Press, 1970.
6. Le Toan, T., Lopes, A., Huet, M., "On the Relationships Between Radar Backscattering Coefficient and Vegetation Canopy Characteristics", *Proceedings of IGARSS'84 Symposium*, Vol. ESA SP-215, August 1984, pp. 155-160.
7. Ulaby, F.T., Moore, R.K. and Fung, A.K., *Microwave Remote Sensing, Active and Passive*, Addison-Wesley Publishing Company, Vol. 1,2 and 3, 1981, 1982, and 1986.

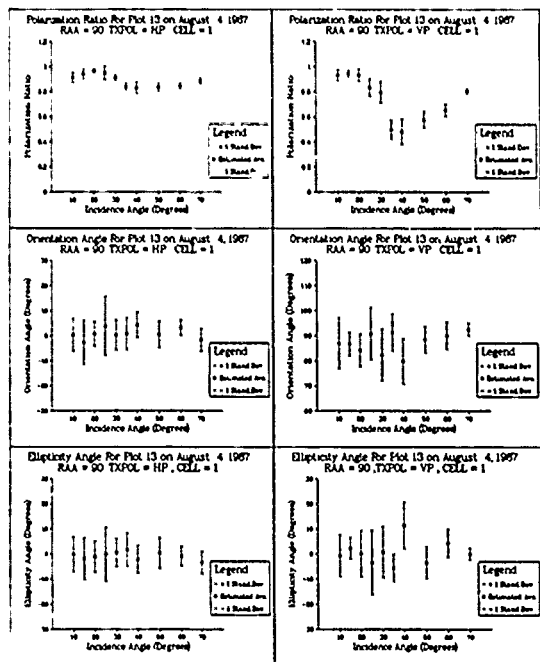


Fig. 2. Polarization State Parameters vs Incidence angle for Wheat(Katepawa) on August 4, 1987, at Row Aspect Angle = 90° , for Range Cell 1

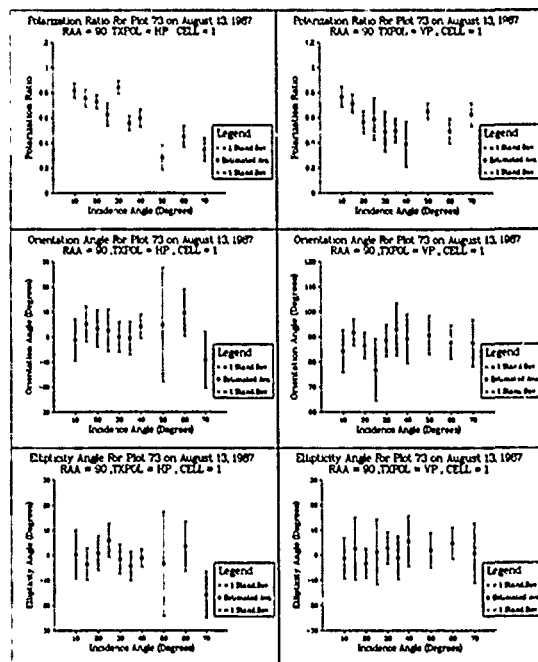


Fig.3: Polarization State Parameters vs Incidence angle for Canola(Tobin) on August13, 1987, at Row Aspect Angle = 90° , for Range Cell 1

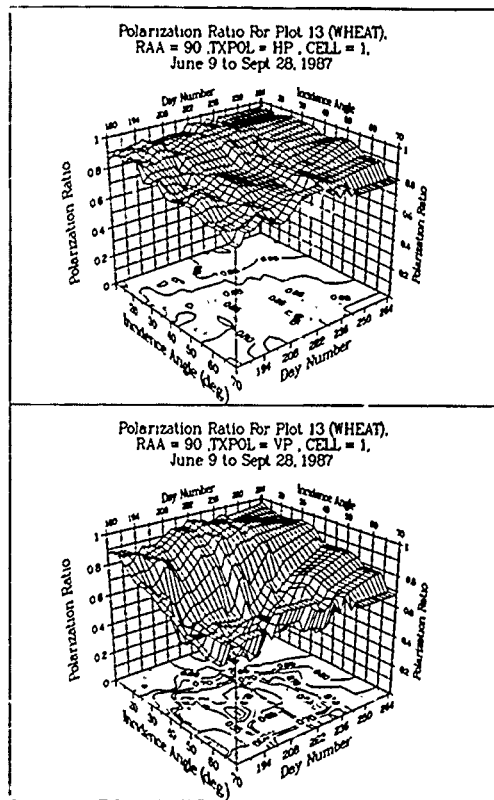


Fig.4. Polarization Ratio vs Incidence Angle and Time for Wheat(Katepawa) at Row Aspect Angle = 90° , at Range Cell 1

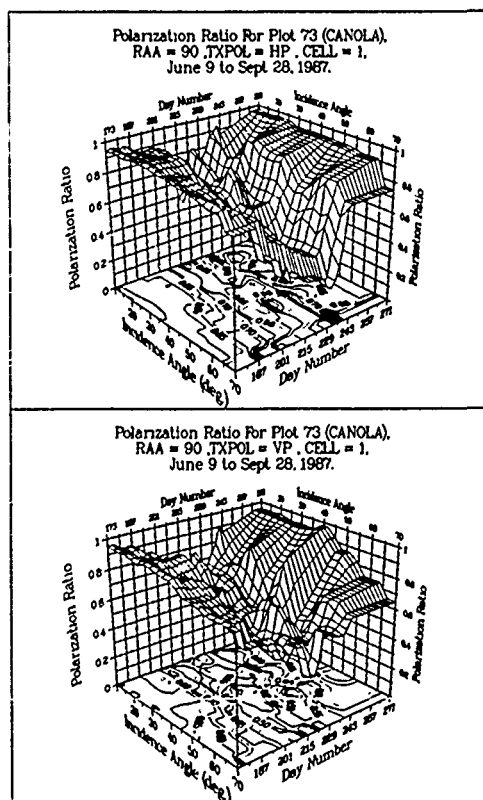


Fig.5: Polarization Ratio vs Incidence Angle and Time for Canola(Tobin), at Row Aspect Angle = 90° , at Range Cell 1

A MICROWAVE BACKSCATTER EXPERIMENT OF A BURDOCK FIELD BY
AN X-BAND FM-CW RADAR

By

Takashi KUROSU, Takeshi SUITZ and Toshihiko UMEHARA

Communications Research Laboratory,

Ministry of Posts and Telecommunications

4-2-1, Nukui-Kitamachi, Koganei-shi, Tokyo 184, JAPAN

Abstract

Classifying crop fields is one of interesting objectives in developing remote microwave sensing techniques.

A backscatter experiment of a burdock field by an X-band FM-CW radar was carried out in order to obtain basic data for this study. That is, the radar was moved 2.3 m at the constant speed of about 7 cm/s in the direction perpendicular to the ridge line with the beam direction parallel to the ridge line at the antenna heights of 3.5 m and 2.0 m above the top of canopy level at the incident angles of 45 degrees and 60 degrees, respectively.

As a result, the ridge interval in the field and the leaf size could be detected by the analysis of the backscattered power fading, e. g., deriving the auto-correlation function. The detection of these sizes could be also found to depend on the incident angle. Moreover, an interesting dependence of the polarizational combination of irradiated and backscattered waves on the burdock plant shape could be found from observation for the four polarizations, VV, VH, HH and HV.

Key words. Microwave Backscattering, Burdock Field, X-band FM-CW Radar, Fading, and Plant Shape

1. Introduction

Surface roughness, furrow direction and plant shape are important parameters in classifying crop fields. A crop field usually has a periodic feature due to a

ridge, namely one of a set of raised strips in the field. Analysis of fading is an attempt to extract the information of plant shape.

The wavelength of periodic feature of

homogeneous medium, e. g., ocean surface can be detected by the auto-correlation function of backscattered power fading data (Weissman and Johnson, 1979). We detected the periodic characteristics, i. e., ridge interval and ear size, of a rice vegetation, which is composed of heterogeneous media, by analyzing fading data (Kurosu et al., 1980). They also found that a periodic fading caused by a periodic characteristic of vegetation can make a two-peak configuration which is superposed in the probability density function of backscattered power fading data.

A backscatter experiment of a burdock field was carried out in Tsukuba city of Ibaraki prefecture in Sep., 1987. Observational and analytical results will supply a information to crop classifying.

2. Experiment

The used radar is an X band FM-CW radar with the operational frequency of 9.6 GHz, with the triangular modulation, with the sweep width of 200 MHz, with the modulation frequency of 850 Hz and with the transmitter power of 10 mW. A circular horn antenna was used, which is co-operational for transmitting and receiving with the beam width of 11 degrees.

The IF output of the receiver with the beat frequency was supplied to the high pass filter of the lower frequency of 6 kHz and to the band pass filter of the band width of 3 kHz. The output of the latter filter was envelop-demodulated and amplified in logarithmic mode. The Log. Amp. output of the backscattered power was

analog-recorded on a magnetic tape and on a chart.

A burdock field (seeding day: May 16) was taken a photograph from the direction parallel to the ridge line at the sight (incident) angle of about 45-degrees, and covered by broad leaves as shown in Fig.

1. The ground truth data is listed in Table 1. The observational conditions are already described.

3. Results and discussion

The fading data against moved distance recorded on magnetic tape was converted from analog to digital at the sampling frequency of 100 Hz through the low pass filter of the upper frequency of 100 Hz. Digitalized fading data is shown in Fig. 2 for each incident angle. One subinterval giving 2048 samples was extracted from each interval of 2.3 m to be analyzed. The radar moving speed was maintained constant as far as possible during the observation for each incident angle and each polarization, but different from each other in each observation. So the length of analyzed subinterval is different from each other and given by the range of the auto-correlation function. The method of analysis should be referred to Kurosu et al., 1988, adding using the deviation of observed probability density bar graph from the value at the middle point of each mesh of bar graph of its approximate curve.

The auto-correlation functions are shown in Fig. 3. The vegetation period is given by the well-auto-correlated wavelength. By comparing this wavelength to the ground truth data, one plant part and its size can

be detected. As a result, the ridge interval could be measured by 60 degree incident angle and HH polarization and the average leaf size of about 27 cm was by 45 degree incident angle and VH polarization. And there exists other incident angle and polarization combination which detected these sizes with weak auto-correlation (listed in Table 3).

Bar graphs of each observed probability density of backscattered power, an approximate curve denoted by the character "A" and a weight function are shown in Fig. 4 for each incident angle and each polarization. Values of parameters of weight functions, expansion coefficients of weighted normalized orthogonal polynomials and approximate errors are listed in Table 2 (a)-(c), respectively. Fig. 5 shows the weighted normalized orthogonal polynomial with maximum absolute expansion coefficient for each incident angle and each polarization. For 45 incident angle and VH polarization for leaf size detection, the bar graph in Fig. 4 (a) has two raising portions at backscattered powers giving two peaks of the curve in Fig. 5 (a). This is found in the rice field experiment for the detection of ridge interval (Sultz et al., 1987). All results are listed in Table 3.

4. Conclusions

As shown in Table 3, the ridge interval and the leaf size could be detected by the auto-correlation function, but the polarizations for the ridge interval detection are different from between each incident angle and this applies to the leaf size

detection. From these results, the microwave backscatter at X-band depends on the plant shape and on the incident angle. It may be an interesting problem to study the polarization dependence on the plant shape.

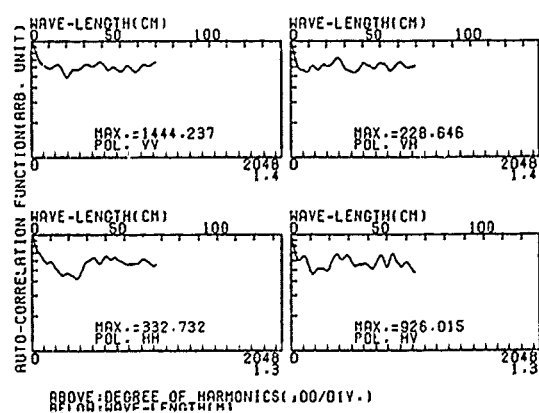
We would express our gratitude to Mr. T. Ojima, the chief of the remote sensing research section and Dr. T. Akiyama, the chief of the Agro-Biological Measurement Lab. of the National Institute of Agro-Environmental Sciences, Ministry of Agriculture, Forestry and Fisheries.

References

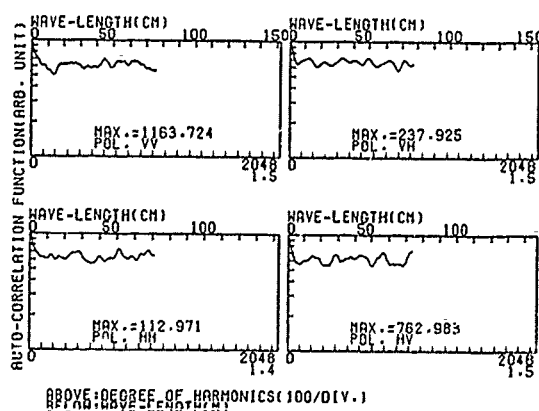
1. Weissman, D. E. and J. W. Johnson, "Rough surface wavelength measurement through self mixing of doppler microwave backscatter", IEEE Trans. Antennas and Propagat., Vol. AP-27, No. 6, pp. 730-737, 1979.
2. Kurosu, T., T. Sultz, T. Koza and T. Umehara, "Detection of periodic characteristics of rice field vegetation by microwave backscatter measurement", J. Comms. Res. Lab., Vol. 35, No. 146, pp. 229-242, 1988.
3. Sultz, T., S. Yoshikado, T. Kurosu, T. Koza and T. Umehara, "Backscattering coefficient of rice crops and rice fields by an X-band scatterometer", Proc. of 21st Symp. on Remote Sens. of Environ., Vol. II, Ann Arbor, Michigan, pp. 767-773, Oct. 1987.



Fig. 1. Photograph of a burdock field.

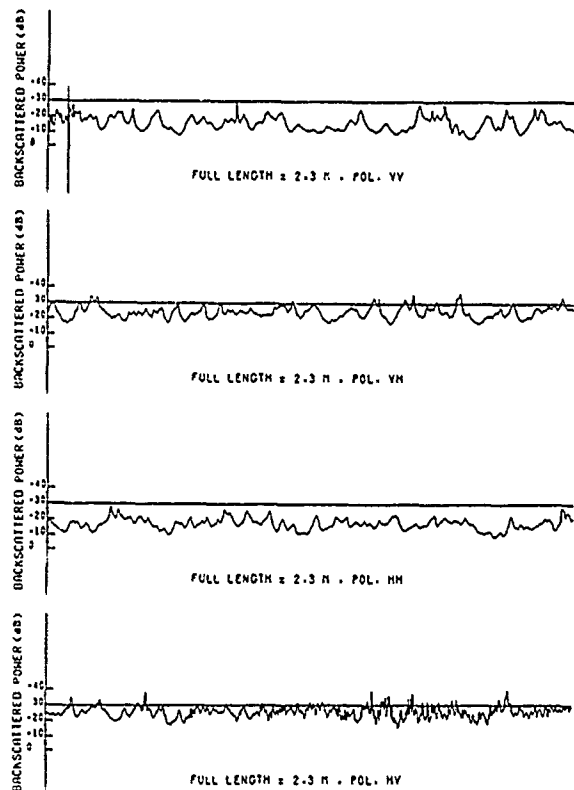


(a) Incident angle: 45 degrees.

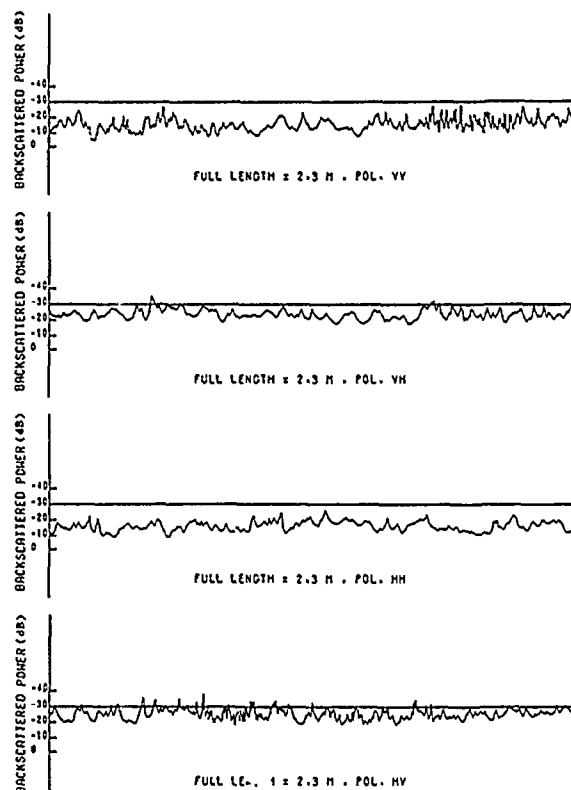


(b) Incident angle: 60 degrees.

Fig. 3. Auto-correlation functions, denoting maximum value and polarization.

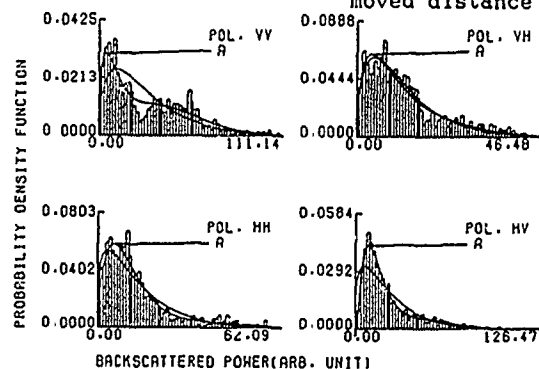


(a) Incident angle: 45 degrees.

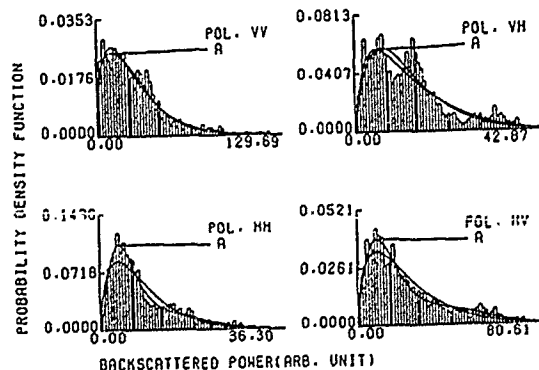


(b) Incident angle: 60 degrees.

Fig. 2. Microwave backscatter data from a burdock field, denoting full length of radar moved distance and polarization.

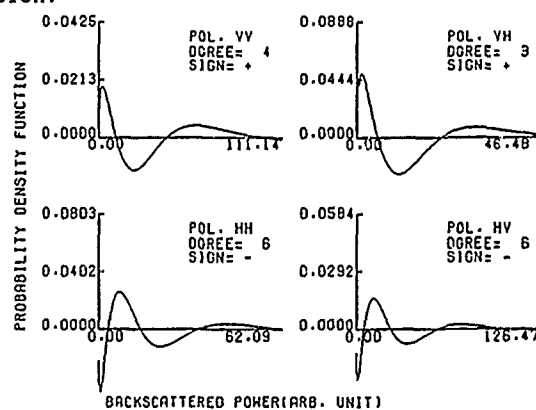


(a) Incident angle: 45 degrees.



(b) Incident angle: 60 degrees.

Fig. 4. Observed probability densities, their approximate curves indicated by "A" and weight functions, denoting polarization.



(a) Incident angle: 45 degrees.

(b) For Incident angle of 60 degrees; omitted.

Fig. 5. Weighted normalized orthogonal polynomials with maximum absolute expansion coefficients, denoting polarization, normalization degree and coefficient sign.

Table 1. The ground truth data of a burdock field.

| Ridge interval | Hill interval | Plant height | Major leaf axis | Minor leaf axis | LAI |
|----------------|---------------|-----------------|-----------------|-----------------|-----|
| 50 cm | 30 cm | ≈ 55 cm | ≈ 30 cm | ≈ 26 cm | 6.6 |

Table 2 (a). Values of parameters of

weight functions.

Incident angle: 45 degrees.

| Parameter | m | s |
|-----------|-------|------|
| Pol. VV | 0.573 | 18.9 |
| VH | 0.603 | 7.40 |
| HH | 0.343 | 10.3 |
| HV | 0.376 | 16.8 |

Incident angle: 60 degrees.

| | | |
|---------|-------|------|
| Pol. VV | 0.646 | 16.3 |
| VH | 1.04 | 6.19 |
| HH | 0.799 | 4.74 |
| HV | 0.648 | 13.2 |

Table 2 (c). Approximate errors (r.m.s. values

($\times 10^2$ (percent density))).

Incident angle: 45 degrees.

| Pol. | VV | VH | HH | HV |
|-------|---------|---------|---------|---------|
| Error | 0.00311 | 0.00615 | 0.00460 | 0.00178 |

Incident angle: 60 degrees.

| | | | | |
|-------|---------|---------|---------|---------|
| Error | 0.00199 | 0.00857 | 0.00543 | 0.00274 |
|-------|---------|---------|---------|---------|

Table 2 (b). Expansion coefficients of weighted normalized orthogonal polynomials.

Incident angle: 45 degrees.

| Degree | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---------|----------|---------|---------|---------|--------|--------|---------|---------|---------|----|
| Pol. VV | 0.158 | 0.212 | 0.169 | 0.0906 | | | | | | |
| VH | 0.0844 | 0.0569 | -0.0115 | -0.0449 | | | | | | |
| HH | 0.0167 | -0.0656 | -0.146 | -0.161 | | | | | | |
| HV | -0.00921 | -0.0861 | -0.147 | -0.163 | -0.149 | -0.123 | -0.0949 | -0.0682 | -0.0430 | |

Incident angle: 60 degrees.

| | | | | | | | | | | |
|---------|---------|---------|---------|--------|--------|---------|---------|---------|--|--|
| Pol. VV | 0.00851 | -0.0229 | -0.0190 | 0.0189 | 0.0535 | 0.0605 | | | | |
| VH | 0.0928 | 0.0654 | 0.00478 | | | | | | | |
| HH | 0.0428 | -0.0216 | -0.0998 | -0.129 | -0.110 | -0.0718 | -0.0375 | -0.0159 | | |
| HV | 0.0955 | 0.0606 | -0.0378 | -0.109 | -0.118 | -0.0794 | | | | |

Table 3. The analytical results of the fading for a burdock field.

Incident angle: 45 degrees.

| Polarization | Item | Maximum absolute expansion coefficient | | Existence of auto-correlation | Correlated distance (cm) | Corresponding plant part |
|--------------|------|--|--------|-------------------------------|--------------------------|--------------------------|
| | | Degree | Value | | | |
| VV | | 4 | 0.212 | none | 50 (inversely)* | ridge |
| VH | | 3 | 0.0844 | exist. | ≈ 27 | leaf |
| | | | | weak exist. | 50 | ridge |
| HH | | 6 | -0.161 | none | | |
| HV | | 6 | -0.163 | none | ≈ 27 (inversely) | leaf |
| | | | | none | 50 (inversely) | ridge |

Incident angle: 60 degrees.

| | | | | | |
|----|---|--------|-------------|----------------|-------|
| VV | 8 | 0.0605 | none | 50 (inversely) | ridge |
| VH | 3 | 0.0928 | weak exist. | 50 | ridge |
| HH | 6 | -0.129 | weak exist. | ≈ 27 | leaf |
| | | | exist. | 50 | ridge |
| HV | 7 | -0.118 | none | 50 (inversely) | ridge |

Note:

*; This means that the auto-correlational value inversely decreases.

RADAR EN AGRICULTURE : RÉSULTATS PRÉLIMINAIRES SUR UN SITE AGRICOLE AU QUÉBEC

K. P. B. Thomson, R. Landry, G. Edwards et M. Cantin

Centre de Géomatique
Département des Sciences Géodésiques et de Télédétection
Université Laval
Ste-Foy, Québec

H. Gwyn

CARTEL
Université de Sherbrooke
Sherbrooke, Québec

ABSTRACT

During the summer of 1988, C and X band SAR data were obtained over an agricultural test site in Québec. The study area is situated in the municipality of Saint-Léonard d'Aston on the south shore of the St. Lawrence river. The SAR data were acquired at four times during the growing season (May 16, June 6, July 15 and August 4). The site was imaged from three different directions on each date. One direction was perpendicular to the field direction. A second direction was flown parallel and a third was approximately 45° to the field direction. Data on the various crops were also acquired during each overflight.

The SAR data for July, at two illumination angles, have been registered to a map base and classified. The agricultural data, the cartographic information and the classification results have integrated into a Pamap geographical information system. The preliminary results show that the illumination direction is important for certain crops. This effect can also influence the overall classification accuracy. The highest overall classification accuracy was obtained with the X band data.

Keywords: SAR, Agriculture, Classification.

RÉSUMÉ

Le but de cette communication est de présenter des résultats préliminaires d'analyse des données ROS (Radar à Ouverture Synthétique) aéroporté sur un site agricole au Québec. Le site à l'étude est à Saint-Léonard d'Aston dans la vallée du Bas Saint-Laurent.

Pendant l'été 1988, quatre survols du ROS aéroporté du Centre canadien de télédétection ont été effectués sur le site. Les dates des survols étaient le 16 mai, le 6 juin, le 15 juillet et le 4 août. À chacune de ces dates, la zone d'étude a été survolée dans trois directions (169°, 232° et 132°), à deux fréquences (la bande C et la bande X) et à une polarisation HH.

L'analyse préliminaire s'est faite à partir des données ROS du mois juillet 1988 à deux angles d'illumination. Ces données ont été corrigées géométriquement à la carte cadastrale du site et ont été stockées avec les données agricoles dans un système d'information à référence spatiale (S.I.R.S.).

L'analyse des résultats de classification démontrent que la direction d'illumination du radar par rapport à l'orientation des champs s'avère importante pour certaines cultures. La bande X donne le plus haut pourcentage de classification pour les cinq classes d'utilisation du sol retenues.

Mots-clés: Radar, Agriculture, Classification.

1. INTRODUCTION

Le but de cette communication est de présenter des résultats préliminaires d'analyse des données ROS (Radar à Ouverture Synthétique) aéroporté sur un site agricole au Québec. La présente étude constitue un projet conjoint entre l'Université Laval, l'Université de Sherbrooke et le Centre canadien de télédétection (C.C.T.). Le site est à Saint-Léonard d'Aston dans la vallée du Bas Saint-Laurent.

Un des objectifs de l'étude est l'évaluation des données ROS pour des fins d'inventaire agricole. Plus particulièrement, cette communication traitera de l'effet de la direction d'illumination du ROS sur la précision des classifications. Il sera aussi discuté des différences dans les classifications agricoles résultant de l'utilisation simple ou combinée des deux fréquences du radar (bande C ou bande X).

2. LE SITE

Le site à l'étude, couvrant une superficie d'environ 9 km carré, se trouve dans la municipalité de Saint-Léonard d'Aston. Cette région est sur la rive sud du Saint-Laurent, à mi-chemin entre Sherbrooke et Québec. Le relief de la région est peu accidenté. Parmi les principales cultures, on trouve le maïs (25%), les céréales (principalement l'orge 12% et l'avoine 10%), le foin (37%), le pâturage (5%) et les pommes de terre (5%) (Bernier *et al.*, 1987). Dans les champs de foin, on trouve la luzerne et le trèfle. Les pâturages sont souvent des friches herbacées.

3. ACQUISITION DES DONNÉES

Durant l'été 1988, quatre survols du ROS aéroporté du Centre canadien de télédétection (Livingston *et al.*, 1987) ont été effectués sur le site d'étude. Ces survols ont eu lieu le 16 mai, le 6 juin, le 15 juillet et le 4 août. À chaque date, la zone d'étude a été survolée dans trois directions, 169°, 232° et 132° respectivement (voir figure 1). De plus, les données ROS dans la bande C et dans la bande X (polarisation HH) ont été prises pour chacune de ces dates. La résolution spatiale est d'environ 6 mètres au sol pour sept vues. Les données image reçues du C.C.T. ont également subi une compression de racine carrée afin de les réduire en format de 8 bits.

Une visite sur le terrain a été faite à chacune des dates de survol dans le but de recueillir l'information pertinente sur les cultures et sur leur stade de développement. L'ensemble de ces données, ainsi que toutes autres données auxiliaires, ont été colligées dans le système d'information à référence spatiale Pamap.

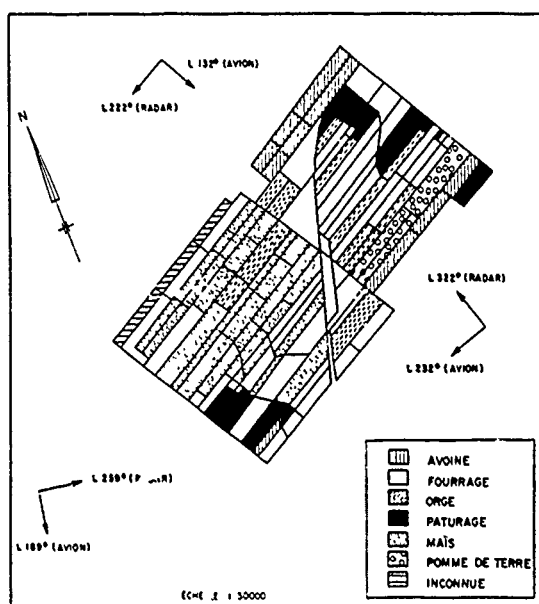


Figure 1: Carte de l'inventaire du site à l'étude avec les azimuts des survols aéroportés.

Pour les fins de cette communication, seules les données de distance oblique prises aux angles d'illumination correspondant aux azimuts de 169° et 232° ont été traitées (voir tableau 1). Elles ont toutes été prises pendant le mois de juillet. Le mois de juillet — le point de départ de l'étude — est le point culminant de la saison végétative dans la région à l'étude. De plus cette époque constitue une période stratégique pour l'acquisition des données en agriculture. L'azimut 232° donne une illumination perpendiculaire à la direction des champs, tandis que celui de 169° donne une direction d'illumination d'environ 45° par rapport aux champs.

| Bande | Direction d'illumination | Polarisation |
|-------|--------------------------|--------------|
| C | 169° | HH |
| C | 232° | HH |
| X | 169° | HH |
| X | 232° | HH |

Tableau 1: Liste des bandes du ROS et azimuts des survols aéroportés

4. MÉTHODOLOGIE

La carte cadastrale de la municipalité (échelle 1:20 000), emmagasinée dans le système d'information à référence spatiale, a servi de carte de base pour la localisation de la zone à l'étude ainsi que de la carte de la vérité de terrain. Cette carte de base a été transférée du système Pamap au système de traitement numérique d'image Perceptron (PCI) où elle a servi comme carte de référence pour la correction géométrique des données ROS. La résolution spatiale pour les deux systèmes requis par cette étude était de 6 mètres.

Suite aux corrections géométriques, une sous-région de 512 lignes par 512 pixels a été sélectionnée. Le choix de cette sous-région a été déterminé par le croisement des trois lignes de vol. Afin d'éliminer le bruit présent dans les images, un filtre

médian 3X3 a été appliqué aux 4 bandes du ROS (Thomson *et al.*, 1987). Finalement, une classification utilisant le critère de maximum de vraisemblance a été effectuée sur les images. Cinq classes de cultures ont été retenues pour les classifications, soit le foin, le maïs, les céréales (orge et avoine), la pomme de terre et le pâturage.

Tous ces résultats de classification ont été transférés et stockés dans le S.I.R.S.

5. RÉSULTATS

Une synthèse des résultats de classification est présentée au tableau 2. Pour certaines cultures, l'effet de l'orientation des champs par rapport à l'illumination du faisceau du radar est important. Ce phénomène est particulièrement présent pour les champs de pommes de terre qui présentent une structure de leurs sillons bien définie. La direction d'illumination perpendiculaire à la direction des champs (232°) donne les meilleurs résultats de classification. Cette influence directionnelle semble aussi importante pour la bande X que la bande C, quoique pour cette dernière, elle soit plus notable.

| Classification | | Pommes de terre | |
|----------------|-------------|-----------------|-------------|
| Bande | % classifié | Bande | % classifié |
| X-232 | 41.66 | X-232 | 69.05 |
| C-232 | 36.93 | C-232 | 43.91 |
| C-169 | 35.60 | C-169 | 23.21 |
| X-169 | 26.57 | X-169 | 12.19 |

| Céréale | | Maïs | |
|---------|-------------|-------|-------------|
| Bande | % classifié | Bande | % classifié |
| C-169 | 34.47 | C-232 | 57.67 |
| X-169 | 26.54 | C-169 | 42.41 |
| X-232 | 24.84 | X-169 | 23.04 |
| C-232 | 18.80 | X-232 | 21.97 |

| Foin | | Pâturage | |
|-------|-------------|----------|-------------|
| Bande | % classifié | Bande | % classifié |
| X-232 | 71.00 | C-232 | 62.04 |
| C-169 | 39.02 | X-169 | 39.14 |
| C-232 | 35.66 | C-169 | 36.59 |
| X-169 | 33.35 | X-232 | 15.26 |

Tableau 2. Synthèse des résultats de classification par maximum de vraisemblance pour cinq classes et les quatre bandes du ROS.

Cette hypothèse est supportée par les valeurs moyennes des signatures provenant des sites d'entraînement. La figure 2 montre que les hautes valeurs pour la bande C-232° sont associées aux cultures de maïs, de céréales et de pommes de terre. Quant à la bande X, la situation est moins évidente. Dans le cas de la bande X-232°, la valeur associée aux céréales est plus élevée que la valeur moyenne obtenue avec la bande X-169° pour une même culture. Par contre, cette même bande donne des valeurs plus basses pour le maïs et la pomme de terre.

La précision globale des classifications des cinq cultures et ce, pour les quatre bandes du ROS, est semblable à celle rapportée

par Vallée *et al.*, (1989) pour la même région en 1987. Malgré des résultats de classification globale peu élevés (environ 40% dans le meilleur cas), il est intéressant de noter que certaines cultures donnent un résultat de classification individuelle intéressant. Par exemple, le foin, le pâturage et les pommes de terre donnent des valeurs de classification de plus de 60%.

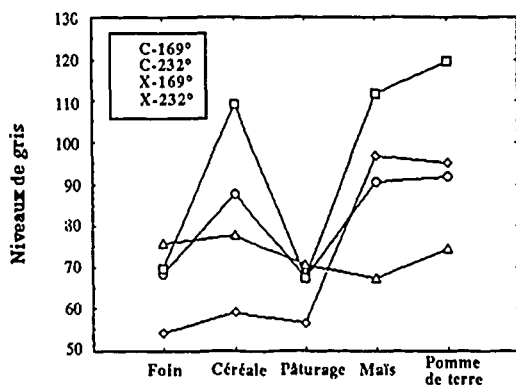


Figure 2. Valeurs moyennes des signatures des sites d'entraînement obtenues pour chaque culture ainsi que pour chacune des quatre bandes du ROS.

Des combinaisons de deux fréquences/un angle et d'une fréquence/deux angles, seule la combinaison des bandes C et X à 232° montre une augmentation significative du pourcentage classifié des cinq cultures (voir tableau 3).

| Classification | |
|----------------|----------|
| Bande | % class. |
| CX-232 | 56.47 |
| C-169/232 | 36.37 |
| CX-169 | 35.44 |
| X-169/232 | 35.17 |

Tableau 3. Synthèse des résultats de classification par maximum de vraisemblance pour cinq cultures (combinaisons des quatre bandes du ROS).

6. CONCLUSIONS

Bien que les résultats de cette communication soient préliminaires, il est possible de tirer certaines conclusions:

1. Dans le contexte de la résolution spatiale examinée (6 m.), l'orientation des champs par rapport à la direction d'illumination du radar influence les résultats des classifications agricoles.
2. Les résultats des classifications obtenus au moyen de la bande X sont légèrement plus élevés que ceux obtenus avec la bande C.
3. Une combinaison des bandes X et C, avec une illumination du ROS perpendiculaire à la direction des champs, donne un résultat de classification au moins 10% plus élevé que le meilleur pourcentage classifié obtenu par des bandes individuelles.

7. REMERCIEMENTS

Les auteurs aimeraient remercier le CRSNG du Canada et les fonds du FCAR du Québec qui ont subventionné cette recherche. L'un des auteurs, R. Landry, bénéficie d'une bourse dans le cadre des actions structurantes du Ministère de l'Enseignement supérieur et de la Science du Québec.

8. RÉFÉRENCES

1. Bernier, M., M. Therrien, et O. Dupont, "Les Données de l'Instrument Thématique (TM) de LANDSAT : des Outils Valables pour la Cartographie du Milieu Agricole du Québec", Comptes-rendus du 10e Symposium Canadien de Télédétection, Edmonton, Alberta, Vol. 2, pp. 661-672, 1986.
2. Livingstone, C. E. *et al.*, "CCRS C-band Airborne Radar—System Description and Test Results", Proceedings of the Eleventh Can. Symp. on Remote Sensing, Waterloo, Ontario, 1989.
3. Thomson, K. P. B., S. Poirier, G. B. Béné, C. Gosselin et G. Rochon, "Filter Selection and Processing Methodology for Synthetic Aperture Radar (SAR) Data in Agricultural Applications", Canadian Journal of Remote Sensing, Vol. 13, No. 1, pp. 6-10, 1987.
4. Vallée, L. B., K. P. B. Thomson, G. B. Béné, "Segmentation et Classification des Cultures", Comptes-rendus du 11e Symposium Canadien de Télédétection, Waterloo, Ontario, pp.147-154, 1989.

**INVENTAIRES AGRICOLES PAR TELEDETECTION:
VALIDITE DES ESTIMATEURS BASES SUR LA MATRICE DE CONFUSION**

- J. Delincé, Projet Agriculture, Institut de Télédétection,
C.C.R., 21020 Ispra, Italie. Telex 380042 EUR 1
Telefax 39/332/789001.
- G. Gérard, Unité de Biométrie et d'Analyse de Données, Faculté
d'Agronomie, U.C.L., 1348 Louvain-la-neuve, Belgique.

L'utilisation des données satellitaires pour améliorer la précision des inventaires agricoles a été tentée avec succès par différents services statistiques dont les offices américains, canadiens, français et italiens. Chaque fois, la fusion des données d'enquête de terrain et des informations satellitaires a été réalisée par la méthode de régression, classique dans le cadre de l'échantillonnage en population finie.

Récemment, deux méthodes (Priesley et Smith 1987, Hay 1988) ont été présentées proposant de corriger l'estimation résultant de la classification numérique des images satellitaires en utilisant l'information contenues dans les matrices de confusion (classement x données terrain). Dans les deux cas, les matrices de probabilités conditionnelles permettent de corriger l'estimation en tenant compte des erreurs dites de commission ou d'omission.

Le but de cet article est de redéfinir les deux estimateurs dans un même système de notation, de spécifier leurs domaines de validité selon le plan d'échantillonnage, et de présenter les formules de calcul des variances asymptotiques.

Priesley S.P. and Smith J.L. (1987) Using Classification Error Matrices to Improve the Accuracy of Weighted Land-cover Models. Photogrametric Engineering and Remote Sensing, vol 53 PP 1259-1263.

Hay A. (1988) The Derivation of Global Estimates from a Confusion Matrix. Remote Sensing Letters.

**SUPERVISED IMAGE INTERPRETATION OF SPOT-DATA FOR THE
DETECTION OF AGRICULTURAL PARCELS IN DIFFERENT VEGETATION
TYPES IN ZAIRE(BANDUNDU).**

Davis L., Goossens R., Ongena T., De Celuwe H.

The aim of the study performed in the region of Bandundu, is to develop a technique to update the agricultural statistics in Zaire. In this study, a number of SPOT-scenes (panchromatic as well as multispectral) are used.

The study includes the following steps:

1. The first stage was to find out which land use classes are important for agricultural statistics.

2. A choice of the test sites. These test sites determined locally, were pointed out because of their differences between landscape and agriculture.

a. LUSANGA: This test site was withdrawn because it was possible to appeal on the agricultural test fields in this village, and also to obtain data, which can be used as a reference for the data collected in the other sites.

b. KATEMBA: The site was withdrawn because a great number of traditional fields are situated in the savanna.

c. MILUNDU: This place was also withdrawn for his traditional fields, situated in a woody region. Many of these fields are situated under a palm tree vegetation.

3. Collecting of ground truth

This includes the following:

a. Collecting of ground control points in order to make a suitable geometric correction possible.

b. Collecting ground data in connection with vegetation and soils.

c. Geometric drawing of the parcels by means of the field survey table.

d. Radiometry

4. To carry out a geometric correction and scale adaptation of the SPOT-images of the three different test sites.

5. Testing of the different image classification techniques which are apparent on the system 600 (1981). In this way, all these different techniques were compared to find out which algorithm was the most suitable for each of these test sites.

6. Further development of the Zaïre project will be emphasized on the improvement of a number of classification algorithms. An optimization of the collecting ground truth and the development of a GIS system in which all geographical information can be stored in such a way that an optimal classification of the agricultural parcels will be possible.

Quantifying Forage and Corn Production with Image Analysis and GIS Techniques

H. Schreier, S. Brown, A. Ryan and R. Wiart

Department of Soil Science, University of British Columbia, Vancouver, B.C. V6T 2A2

Abstract:

Good relationships were obtained between pixel brightness values from digitized colour-IR images, forage biomass and corn quality in six agricultural fields. In the corn fields significant correlations were obtained between crude protein content, foliar phosphorus and pixel brightness values, but the strength of these relationships varied from field to field. This suggests that inherent soil and site conditions influence such analysis. By incorporating the remote sensing, soil and site data into a GIS system we have improved the assessment and obtained crop quality maps which combine crude protein and foliar phosphorus evaluations with site conditions. Economic data can now be incorporated into this data set and the cost of production and the causes of crop performance variations can be examined interactively.

Key words: image analysis, GIS, crop production, crop quality.

Background:

Crop quality and quantity are critical parameters in most agricultural research, yet conventional field determinations of crop performance are tedious, time consuming and difficult to make on a real time basis. Many authors have reported significant relationships between biomass production and near infrared reflection, but such reflection appears to be site and crop specific (Plummer 1988). Variables such as vegetation structure, soil background, moisture content, topographic conditions, live/dead biomass, shadow, and litter influence such relationships. To partition the influence of these variables on the overall reflection is not a simple task (Huete and Jackson 1987; Wilson and Tueller 1987). Some success has been reported in determining foliar nitrogen (Hinzman et al. 1986; Plummer 1988) and foliar phosphorus (Vickery et al. 1980) from remote sensing data. However, the above mentioned problems remain a significant drawback.

In order to improve such a methodology we examined the variability of dry biomass and crop quality in four corn and two forage fields in Matsqui British Columbia. Multi-date digitized colour IR imagery was then combined with other soil and site variables using micro-computer based GIS techniques.

Methods:

Between 20 and 100 neutron tubes were placed in each field and the soil moisture was monitored

over the entire growing season to a depth of 100 cm. The soils were sampled at each of the monitored sites and the samples were analyzed for soil nutrient content, bulk density, and available water storage capacity. A detailed topographic survey was carried out in each field using a surface grid density of 5m and a height accuracy of 5 cm. The forage crop was harvested five times over the growing season and the corn crop at the end of the season. Dry biomass and foliar nutrient content was determined in all samples. 23 x 23 cm format colour-infrared images were obtained before each harvest using an RC-10 Photogrammetric camera and the images were scanned at 50 um resolution using an Optronics scanner. The images were displayed on a PC-based image analyzer and the individual markers at each test site could readily be identified. The ground resolution of each pixel represented 0.2 x 0.2 meters. The pixel brightness values for the three dye layers were extracted and used for statistical analysis with the field and laboratory data set. The data was incorporated into a Geographic Information System (GIS). Biomass and crop quality data was combined with soil data to arrive at crop quality and soil quality combination maps.

Results:

The fields are located in the most productive alluvial floodplain in B.C. and have a ridge and depression topography which ranges between 1-4 meters. The ridges are slightly sandier than the depression and in the latter part of the season some moisture stress was apparent. Good relationships were obtained between crop biomass and selective foliar nutrients in both of the two forage fields. The most significant correlations presented in figure 1 indicate that water balance has an influence on both biomass production and forage quality (foliar phosphorus, and foliar nitrogen).

Pixel brightness values were related to dry biomass and foliar nutrients in the latter part of the season. Total dry biomass could be predicted from the remote sensing data using the following multiple regression equation:

$$DM = -8.3 + 0.1 G - 0.2 R + 0.07 IR$$

$$R^2 = 0.68, SE = 0.45$$

where G = Pixel brightness value of green sensitive dye layer,
 R = Pixel brightness value of red sensitive dye layer,
 IR = Pixel brightness value of IR sensitive dye layer.

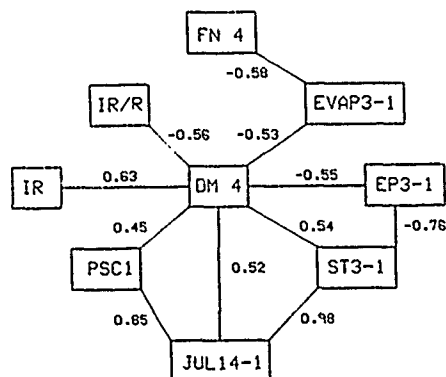


Figure 1. Correlogram for forage yield (dry matter) cut 4, 1987.

The pixel brightness data of the entire image was then reduced to a 2 x 2 m pixel resolution using an affine program and the resulting data set was used to predict the dry matter production of the entire field. The results were then displayed with the GIS technique providing a forage crop yield map for the field (Figure 2).

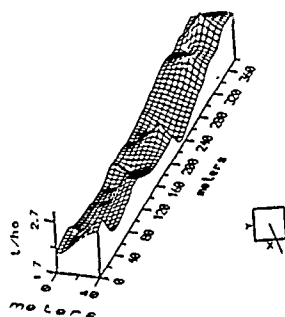


Figure 2. Forage yield cut 3, 1987, predicted from pixel brightness.

The same techniques were used to assess the corn production in the four other test fields. A number of significant relationships were obtained between soil and crop variables and as shown in Figure 3, the IR pixel brightness values were significantly related to elevation, available water storage capacity, crude protein, and foliar phosphorus content in the corn. Relationships between corn biomass and remote sensing data was not consistent. In one field there was a weak relationship between pixel brightness values and cob weight, while stalk weight was significantly related to pixel brightness values in a second field. No remote sensing versus biomass relationships were found in the other fields. However, the relationships between foliar phosphorus and pixel values and between crude protein content and pixel values were more consistent. The later relationship was significant in all four fields while the former was significant in three out of four fields.

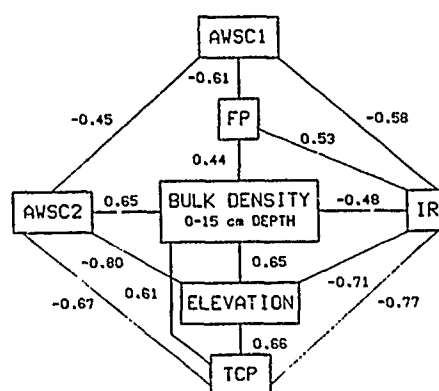


Figure 3. Significant correlations between soil, site and corn crop variables.

The strength of these correlations varied between fields and so did the predictive accuracies. Multiple regressions explained between 37 - 60% of the total variance in foliar phosphorus, and between 24 - 59% of crude protein content. A further improvement of 5 - 10% in R^2 values was obtained when selective soil variables were included in the multiple regression equations. Remote sensing data alone is not entirely satisfactory in predicting corn quality conditions. Since corn quality appears to be related to both remote sensing and elevation, the incorporation of the remote sensing and elevation data with the GIS technique is desirable.

As a first step in using the GIS technique, a crop quality map was produced for one of the corn fields where the relationships between the remote sensing data and the crop quality variables were highest. First a crude protein map was produced by converting the pixel values into crude protein content values using the best regression equation. This resulted in a protein variability map for the field and with the GIS technique the field was then divided into three categories. Class one represented corn with crude protein content below minimum requirements for animal feed. Class two reflected values above minimum feed requirements but below one standard deviation above the mean. Finally class three reflected corn with high crude protein content with values greater than one standard deviation above the mean. A similar map was produced for foliar phosphorus. The resulting classifications were then combined with the GIS technique to produce a protein/phosphorus combination map. The 3 x 3 matrix could successfully be reduced to 4 instead of the theoretically 9 classes, as some combinations did not exist.

The four crop quality classes were found to be significantly different from one another using a Mann Whitney significance test. The intended use of this classification is two fold: 1. Selective crop harvesting according to crop quality for computerized feeding in dairy farms, and 2. For evaluating possible causes of crop quality differences in the field. This latter task can be accomplished by displaying soil and site conditions with the GIS technique in relation to the crop quality.

The inherent soil conditions appear to have a direct influence on the crop quality. The pixel values from the bare field (obtained at the beginning of the growing season prior to planting

of the corn) were related to the corn yield and quality data obtained at the end of the season. In two of the four fields these relationships were significant and this suggests that differences in the soil pattern has an impact on the crop performance. In both fields, the correlations between pixel values and crop quality were between 10 and 20% higher than between the soil pixel values and the crop quality.

The GIS may then be used to determine if the differences in crop quality is large enough to merit selective field management and harvesting. This can be done by assigning a cost of production value to the various yields obtained in the different parts of the field. An example of such a crop production cost map is provided in Figure 4 and reflects the results of production costs in one of the forage fields (Smith 1988).

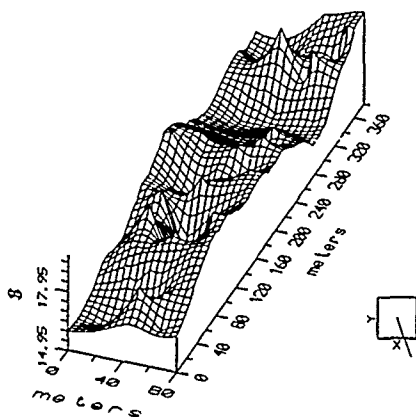


Figure 4. Cost surface of dryland forage production for cut 1, 1987.

Conclusions:

The results from this study indicate that good relationships were obtained between pixel brightness values from digitized colour IR aerial photographs and crop variables. In forage fields with high soil variability, biomass and digestible energy were significantly related to pixel brightness values. In corn fields, pixel brightness/biomass relationships were poor and inconsistent. However, crude protein and foliar phosphorus was highly and consistently correlated with pixel brightness values. The strength of these relationships varied between the four fields suggesting that each field has unique conditions that influence crop quality.

These assessments can be improved if the remote sensing data is incorporated into a GIS system, where the remote sensing data set can be combined with soil and site variables. Crude protein and foliar phosphorus variability were predicted from the remote sensing data set and classifying and combining the results with GIS techniques resulted in the production of a four class crop quality map for the field. The GIS technique can now be used to examine possible causes for the within field crop quality variation by displaying the soil and site conditions for each class and determining whether the differences are significant.

The cost of production in relation to yield and quality can then be displayed spatially and can be used to determine if the crop performance differences are large enough to merit selective field management and harvesting.

We envision that such micro-computer based analysis will assist field managers in considering selective field management and harvesting and this will bring automated dairy farm management a step closer.

References:

- Hinzman, L.D., Bauer, M.E., and Daughtry, C.S.T., Effects of nitrogen fertilization on growth and reflectance characteristics of winter wheat. *Remote Sens. Environ.*, Vol. 19, pp 47-61, 1986.
- Huete, A.R., Soil-dependent spectral response in a developing plant canopy. *Agron. J.*, Vol. 79, pp 61-68, 1987.
- Huete, A.R., and Jackson, R.D., Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands. *Rem. Sens. Environ.*, Vol. 23, pp 213-232, 1987.
- Plummer, S.E., Exploring the relationships between leaf nitrogen content, biomass and the near-infrared/red reflectance ratio. *Intern. J. Rem. Sens.*, Vol. 9, pp 177-183, 1988.
- Smith, S.M., Assessing variability in the production of pastures using GIS with remote sensing techniques. Ph.D. Dissertation, University of British Columbia, Vancouver, B.C., pp 264, 1988.
- Wilson, O.R., and Tueller, P.T., Aerial and ground spectral characteristics of rangeland plant communities in Nevada. *Rem. Sens. Environ.*, Vol. 23, pp 177-191, 1987.
- Vickery, L.J., Hedges, D.A., and Duggin, M.J., Assessment of fertilizer requirements of improved pasture from remote sensing information. *Rem. Sens. Environ.*, Vol. 9, pp 131-148, 1980.

The Effects of Free Canopy Water on SAR Crop Separability

Brisco, B.¹, R.J. Brown¹, and T.J. Pultz²

¹Canada Centre for Remote Sensing
2464 Sheffield Road, Ottawa, Ontario K1A 0Y7

²Intera Technologies Ltd.
1525 Carling Avenue, Suite 600, Ottawa, Ontario K1Z 8R9

ABSTRACT

Several SAR data sets acquired by the Canada Centre for Remote Sensing between 1984 and 1988 were used to evaluate the effects of free canopy water on crop separability. The test sites imaged were located in Saskatchewan, Canada with canola, spring wheat, barley, fallow, and pea crop types present. A separability analysis of these data using divergence statistics indicated that wheat separability may be increased when the crop canopies contain free water. This phenomenon may also be enhanced using VV polarization. These results are very preliminary but have important implications if a consistent relationship can be established.

Keywords: SAR, Crop Identification, Canopy Water

1. INTRODUCTION

The timeliness of data availability for crop information systems is of utmost importance to yield estimation. For remote sensing techniques to provide a valuable input the throughput time for data acquisition and analysis needs to be less than that using conventional survey approaches. Synthetic aperture radar (SAR) systems provide all weather, day or night operation and thus can help fulfill this requirement.

Numerous investigations have demonstrated the potential of SAR data for crop identification, with a recent summary available in Ulaby et al., (1986). However, most SAR data analysed for this purpose has been acquired under dry canopy conditions. Ground-based microwave measurements show an increased radar backscattering coefficient for wet canopy conditions (Allen and Ulaby, 1984; Sofko et al., 1989). These measurements have been made for a limited number of crop types and conditions because of the logistical difficulties associated with this type of data acquisition. Due to the above mentioned 'all weather'

capabilities of SAR systems more 'wet' data sets will become available as the planned SAR satellites are launched in the 90's.

The Canada Centre for Remote Sensing (CCRS) has obtained several SAR data sets of agricultural test sites in Saskatchewan, Canada which were acquired when the crop canopies contained free water. These data were uncalibrated and from different sensors and subsequently were not suited to detailed quantitative analysis. However, their uniqueness warranted investigation and an analysis plan using divergence statistics as a measure of crop separability was developed. This paper presents the results of this investigation.

2. SAR DATA DESCRIPTION AND ANALYSES

On August 6, 1986 the Intera Technologies Ltd., STAR-1 system acquired X-HH SAR data over several test sites in Saskatchewan, Canada (Figure 1). At two of the sites, Outlook and Melfort, it was raining (or had just rained) at the time of data collection. Consequently, the crop canopies were wet at the time of imaging. For comparative purposes X-HH data acquired over the same two test sites on August 5, 1984, by the CCRS Convair SAR-580 system were used. The crop canopies were completely dry when this data was acquired. On June 21, 1988 the CCRS X/C SAR acquired C-HH and C-VV data over the Melfort test site just after an extensive rain. As a result the crop canopies once again contained free water at the time of imaging.

A brief description of these three SAR systems can be found in Table 1. All of the SAR data used in this analysis are uncalibrated and caution must be used when comparing results from the different systems. To avoid system effects in the data interpretation a similar pre-processing regime was applied to each data set. This resulted in an amplitude domain output product with a 12.5 meter by 12.5 meter pixel, representing an average of approximately 60 looks per pixel for each SAR data set.

The Outlook and Melfort sites are frequently used by CCRS for agronomic remote sensing research projects and consequently detailed information on crop type and condition is available for each date of SAR acquisition. The ground-truth program conducted in support of these projects is described in detail by King and Mack (1984). For the 1984/1986 comparison plant height, growth stage, crop variety, and plant density were used to identify similar fields of spring wheat, barley, canola, peas, and summerfallow. The Melfort test site also had enough fields of flax in 1984 and 1986 to include it as another crop type. Divergence was then calculated on a class basis (by using a theme mask) and the results compared to evaluate crop separability on the "wet" versus dry" SAR data. The use of the divergence statistic also facilitates inter-SAR comparison as the mean to standard deviation ratio and not the absolute values are the inputs for determining separability.

For the June 21, 1988 Melfort data only growth stage was used to select similar fields of the same crop types as above. This was done in the western portion of the test site because an east/west split in crop condition (due to seasonal rainfall) was apparent from ground data and farmer interviews. Once again divergence was calculated on a class basis and the separability of HH polarization compared to VV polarization. The number of fields and pixels in each class for each SAR data set are given in Table 2. All image analysis and pre-processing of these data were done using an ARIES image analysis system and the CCRS Landsat Digital Image Analysis System (LDIAS).

3. RESULTS AND DISCUSSION

Due to the page length restrictions in these proceedings the images will not be presented in the paper but will be available for inspection during the poster presentation. For the 1984 "dry" and "wet" SAR data of Melfort the divergence statistics calculated are presented in Table 3. One can see that for the "dry" data (1984) canola is the most separable class. This agrees with previous studies which concluded that the broad-leaved canola plant exhibits high backscatter values early in the growing season and thus is quite separable from the other crops grown in the Canadian prairies (Brisco et al., 1989, Brown et al., 1984). For the "wet" data (1986) note that the wheat class becomes readily separable from both canola and flax and that overall the separability of the grain classes increases while that of the other classes decreases.

Divergence statistics calculated from the dry and "wet" SAR data for the Outlook site are presented in Table 4. For the "dry" data from Outlook the summerfallow category is readily separable. By August summerfallow fields have become quite smooth and dry and thus have low backscattering coefficients and are readily discriminated. This result has also been previously reported (Brisco et al., 1988, Brown et al., 1984, Cihlar et al., 1984). However, the "wet" data (1986) divergence statistics indicate it is the wheat category which is readily separable. For the studies cited above the wheat class was generally the most poorly identified because of its overlapping tonal signature with most other crop types, especially barley.

For both the Outlook and the Melfort sites the wheat class separability increased under wet canopy conditions. The authors attribute this effect to the

lack of plant parts in a wheat canopy which collect and hold water, when compared to the other crops investigated. Even the barley plant, which is very similar in structure to the wheat plant, has a larger auricle at the stem-leaf junction which can serve as a water trap (Figure 2). This may account for the increase in wheat-barley separability in the Outlook "wet" SAR data although the wheat-barley separability decreased for the Melfort "wet" data. The wheat canopy holds less water and dries out faster after a rainfall than other crop canopies. Thus wheat fields become the darkest agricultural areas on "wet" SAR data. This may allow the structural differences in the crops, particularly between wheat and barley, to be enhanced in the microwave regime by acquiring "wet" data.

The divergence statistics for the "wet" Melfort C-HH and C-VV data from June 21, 1988 can be found in Table 5. Note that this is C-band data and no comparable "dry" data set is used for comparison. These data were included in the analysis because they represented another "wet" data set and because anomalous assessment with quick-look data in the field indicated differences between the VV and HH data. The horizontally polarized data exhibits generally weak class separability, while the vertically polarized data shows an increased separability for wheat (from canola and peas) and barley (from peas). The other classes have similar or lower separability for the two polarizations.

Several authors have discussed the vertical orientation of grain canopies with respect to HH and VV attenuation (Allen and Ulaby, 1984; Ulaby and Wilson, 1985; L. Toan et al., 1984). In general, attenuation is found to be more significant for VV polarization than HH polarization. Consequently, the HH polarization has greater penetration and the effects of free water in the canopy are not as pronounced for this polarization.

For all "wet" data sets the images appeared to have less contrast (ie class means were closer together) than for "dry" images. This would tend to reduce the overall classification accuracy due to increased confusion between the various crop classes. However, if the wheat crop classification is improved under "wet" conditions such SAR data would be a useful addition to crop discrimination techniques.

4. CONCLUDING REMARKS

Several SAR data sets that were acquired when free water was present in the crop canopy were examined to determine the utility of these data for crop separability purposes. An examination of divergence statistics for spring wheat, barley, pea, canola, and summerfallow crop types indicated that "wet" data may enhance separability within the grain crops. This result is related to the canopy structure and may offer an important mechanism for improving crop classification with radar data.

While these results are preliminary their significance is considerable if the response described above is consistent. Future experiments with both the CCRS SAR and ground-based scatterometers are planned to further investigate this phenomenon. These experiments will acquire calibrated data from comparable wet and dry canopy conditions to overcome some of the limitations of the data used in this

analysis. If these results are repeatable the technique could be used to improve grain crop classification accuracy by using SAR data acquired when the canopies are wet. These "windows" of opportunity could be identified by using existing meteorological observation technology.

5. ACKNOWLEDGEMENTS

Thanks to Ms. G. Allen for typing the manuscript and Ms. M. Trindade for figure, poster, and slide preparation. Thanks also go to Mike Manore for critically reviewing the manuscript and Ms. C. Huett for help with the image analysis.

6. LITERATURE CITED

- Allen, C.T. and F.T. Ulaby. "Characterization of the microwave extinction properties of vegetation canopies". Radiation Laboratory, Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109, 1984a.
- Allen, C.T. and F.T. Ulaby. "Modelling the polarization dependence of the attenuation in vegetation canopies". Proceedings of IGARSS'84 Symposium, Strasbourg, 27-30 August, 1984b.
- Brisco, B., et al., "Early season crop discrimination with combined SAR and TM Data", Canadian Journal of Remote Sensing, Volume 15, No. 1., 1989.
- Brown, R.J., et al., "Crop type determination from multi-temporal SAR imagery", 9th Canadian Symposium on Remote Sensing, St. John's, Newfoundland, 14-17 August, 1984.
- Cihlar, J., and T. Hirose. "SAR response of agricultural targets in a northern prairie environment", Proceedings of IGARSS'84 Symposium Strasbourg, France, 27-30 August, 1984.
- King, G.J., and A.R. Mack. "A manual for recording significant agricultural ground information in remote sensing programs". Agriculture Canada, Ottawa, 1984.
- Le Toan, T., et al., "On the relationship between radar backscattering coefficient and vegetation canopy characteristics", Proceedings of IGARSS'84 Symposium, Strasbourg, France, 27-30 August, 1984.
- Sofko, G.J., et al., "Variation of microwave radar cross-section of wheat during the initial hours of a rainfall". These proceedings, 1989.
- Ulaby, F.T., et al., "Microwave remote sensing: Active and passive", Volume 3, Artech House, Dedham, Mass., 1986.
- Ulaby, F.T., and E.A. Wilson. "Microwave attenuation properties of vegetation canopies", IEEE Trans. Geos. and Remote Sensing, Volume GE-23, No. 5, September, 1985.

Table 1. System parameters of the three SAR systems for the data acquisitions described in this paper.

| | SAR-580 | Star 1 | CCRS C |
|--------------------------|---------|--------|--------|
| Frequency (GHz) | 9.45 | 9.25 | 5.3 |
| Wavelength (cm) | 3.0 | 3.2 | 7.66 |
| Incidence angle (centre) | 45 | 67 | 65 |
| Polarization | HH | HH | HH/VV |
| Swath Width (km) | 5.2 | 22.4 | 22 |
| Azimuth Resolution (m) | 3 | 4.2 | 6 |
| Range Resolution (m) | 3 | 6 | 6 |

Table 2. The number of fields and pixels of each crop type extracted from each SAR image.

| | | SAR IMAGE | | | | |
|--------------|---------|-----------|------|---------|-------|---------|
| | | Outlook | | Melfort | | Melfort |
| | | 1984 | 1986 | 1984 | 1986 | 1988 |
| Wheat | #Fields | 2 | 8 | 10 | 6 | 15 |
| | #Pixels | 6362 | 2512 | 11005 | 5264 | 23786 |
| Barley | #Fields | 1 | 1 | 9 | 5 | 8 |
| | #Pixels | 2122 | 2147 | 12789 | 6501 | 14625 |
| Canola | #Fields | 3 | 3 | 4 | 9 | 11 |
| | #Pixels | 5028 | 715 | 4652 | 12307 | 18051 |
| Peas | #Fields | - | - | 3 | 5 | 11 |
| | #Pixels | - | - | 2306 | 8150 | 17345 |
| Summerfallow | #Fields | 8 | 6 | 6 | 5 | 7 |
| | #Pixels | 3224 | 4396 | 6564 | 4217 | 7311 |
| Flax | #Fields | 5 | 3 | 3 | 5 | - |
| | #Pixels | 3400 | 2565 | 2550 | 5717 | - |

Table 3. Divergence statistics calculated from August 1984/1986 X-HH SAR data of Melfort, Saskatchewan.

| 1984 "Dry" | | | | | | |
|--------------|-------|--------|--------|------|------|--------------|
| | Wheat | Barley | Canola | Peas | Flax | Summerfallow |
| Wheat | 0.00 | - | - | - | - | - |
| Barley | 0.27 | 0.00 | - | - | - | - |
| Canola | 0.38 | 1.27 | 0.00 | - | - | - |
| Peas | 0.03 | 0.30 | 0.32 | 0.00 | - | - |
| Flax | 0.43 | 0.27 | 1.10 | 0.27 | 0.00 | - |
| Summerfallow | 0.13 | 0.02 | 0.94 | 0.17 | 0.27 | 0.00 |

| 1986 "Wet" | | | | | | |
|--------------|-------|--------|--------|------|------|--------------|
| | Wheat | Barley | Canola | Peas | Flax | Summerfallow |
| Wheat | 0.00 | - | - | - | - | - |
| Barley | 0.03 | 0.00 | - | - | - | - |
| Canola | 0.72 | 0.50 | 0.00 | - | - | - |
| Peas | 0.47 | 0.32 | 0.03 | 0.00 | - | - |
| Flax | 0.86 | 0.61 | 0.01 | 0.07 | 0.00 | - |
| Summerfallow | 0.49 | 0.32 | 0.02 | 0.00 | 0.05 | 0.00 |

Table 4. Divergence statistics calculated from August 1984/1986 X-HH SAR data of Outlook, Saskatchewan.

| 1984 "Dry" | | | | | |
|--------------|-------|--------|--------|------|--------------|
| | Wheat | Barley | Canola | Flax | Summerfallow |
| Wheat | 0.00 | - | - | - | - |
| Barley | 0.04 | 0.00 | - | - | - |
| Canola | 0.34 | 0.41 | 0.00 | - | - |
| Flax | 0.45 | 0.61 | 0.04 | 0.00 | - |
| Summerfallow | 1.18 | 1.30 | 2.94 | 3.17 | 0.00 |

| 1986 "Wet" | | | | | |
|--------------|-------|--------|--------|------|--------------|
| | Wheat | Barley | Canola | Flax | Summerfallow |
| Wheat | 0.00 | - | - | - | - |
| Barley | 1.09 | 0.00 | - | - | - |
| Canola | 2.37 | 0.33 | 0.00 | - | - |
| Flax | 2.92 | 0.46 | 0.03 | 0.00 | - |
| Summerfallow | 2.92 | 0.48 | 0.02 | 0.00 | 0.00 |

Table 5. Divergence statistics calculated from June, 1988 Wet C-HH/VV SAR data of Melfort, Saskatchewan.

| HH | | | | | |
|--------|-------|--------|--------|------|--------|
| | Wheat | Barley | Canola | Peas | Fallow |
| Wheat | 0.00 | - | - | - | - |
| Barley | 0.05 | 0.00 | - | - | - |
| Canola | 0.41 | 0.19 | 0.00 | - | - |
| Peas | 0.57 | 0.28 | 0.05 | 0.00 | - |
| Fallow | 0.19 | 0.19 | 0.29 | 0.58 | 0.00 |

| VV | | | | | |
|--------|-------|--------|--------|------|--------|
| | Wheat | Barley | Canola | Peas | Fallow |
| Wheat | 0.00 | - | - | - | - |
| Barley | 0.17 | 0.00 | - | - | - |
| Canola | 1.21 | 0.58 | 0.00 | - | - |
| Peas | 1.62 | 0.83 | 0.03 | 0.00 | - |
| Fallow | 0.59 | 0.26 | 0.12 | 0.26 | 0.00 |

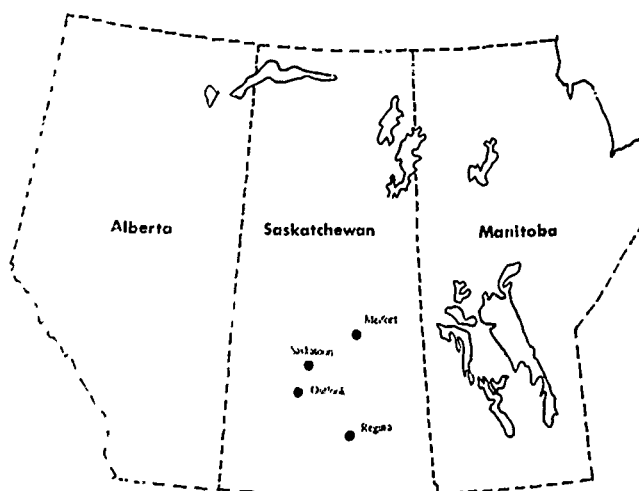


Figure 1. The location of Melfort and Outlook agricultural test sites in Saskatchewan, Canada



Figure 2. Portion of barley plant showing the location of the auricle

ANALYSIS OF C-BAND SAR DATA IN AN AGRICULTURAL CONTEXT

N. Prout, J. Gairns, C. Hutton, D. Mullins
INTERA Technologies Ltd.
1525 Carling Avenue, Suite 600
Ottawa, Ontario, Canada, K1Z 8R9
Telephone: (613) 728-6111
FAX: (613) 728-4009

A. Mack
Land Resource Research Centre
Agriculture Canada
Central Experimental Farm, Bldg. 74
Ottawa, Ontario
K1A 0C6
Telephone: (613) 995-5011

Synthetic aperture radar data (SAR) is being proposed as a spaceborne tool for the 1990's which could be used in routine assessment of agricultural conditions and crop inventory. The current lack of SAR satellites precludes direct evaluation of the utility for such application; however, airborne SAR data for an agricultural site is analyzed in order to examine the potential for crop separation, field boundary detection and the requirement for integration with visible and infrared satellite data.

The mixed crop test site in Manitoba is analyzed with single C-band, multiple-band airborne SAR and LANDSAT data. Extensive ground truth records are used in the assessment of crop separation and field boundary delineation. A satellite simulation of C-SAR is also produced to proposed RADARSAT specifications in order to more fully examine the potential of these data.

PRELIMINARY EVALUATION OF MULTI-DATE SAR DATA
FOR THE IDENTIFICATION OF AGRICULTURAL CROPS IN SOUTHERN ONTARIO

J. A. Fischer and R. S. Mussakowski

Ontario Centre for Remote Sensing
Ministry of Natural Resources
90 Sheppard Avenue East, 4th Floor
North York, Ontario, M2N 3A1

ABSTRACT

Digital airborne SAR imagery was recorded over an agricultural area in Southern Ontario on five dates during the summer of 1988. Mean values of radar return were derived from the five data sets for each of the major field crops grown in the area (corn, soya beans, tobacco and winter wheat). This paper presents graphs of the values obtained from narrow-swath, 6-metre-resolution data in both C and X bands and both HH and VV polarizations. The combination of band, polarization and date found best for identifying each crop is noted.

KEY WORDS: synthetic aperture radar, agriculture, crop inventory, multi-band, multi-date

INTRODUCTION

There is currently much interest in the utility of radar imagery for crop classification because the data could be acquired at any time found to be best for the crop classification, regardless of cloud-cover conditions. The investigation reported in this paper was conducted under the Federal-Provincial Radar Data Development Program of the RADARSAT Project Office, Canada Centre for Remote Sensing (CCRS), as part of a national exercise in the development of agricultural applications of radar data in preparation for the expected launch of RADARSAT in the middle 1990s. The SAR data evaluation was performed at the Ontario Centre for Remote Sensing. Related studies are also underway at the University of Guelph, the University of Waterloo and CCRS.

STUDY AREA

The study area is located in the southeast section of Oxford county, Ontario (Figure 1). The area was chosen for its wide variety of crops and soils. Approximately seventy percent of the area is covered with corn, soya bean, winter wheat and hay/pasture crops growing on loam to clay loam glacio-lacustrine soils. The remaining thirty percent of the area is covered with tobacco and rye crops growing on calcareous loamy to sandy soils (Wicklund and Richards, 1961).

DATA

Imagery for the current study was acquired over a 150-sq.km. (15,000-ha) area of prime agricultural land in the southeast section of Oxford County. SAR imagery was recorded for the area during the 1988 growing season, on May 3, June 2, June 7, July 18 and August 3. Both C and X band imagery were acquired, with HH and VV polarization, in three swath-width modes -- nadir, wide, and narrow. Further processing of the data will be carried out to provide cross-polarized data sets.

To date, only the narrow-mode imagery has been analyzed. This imagery has a swath width of 18 km and a range and azimuth resolution of 6 m. The incidence angle range of the portion of the swath studied is between 60 and 62 degrees.

Colour infrared aerial photographs at a scale of 1:10,000 were also acquired for the entire study area. A detailed ground survey of approximately 20 percent of the area was conducted within two days of each radar mission. A general crop identification survey of the remaining 80 percent of the area was conducted near the end of the growing season. The results presented in this paper were obtained only from the area surveyed in detail.

METHOD

Image Data Processing

The narrow-mode radar data, which had a resolution (i.e., a field of view) of approximately 6 m and a pixel spacing of 3 to 4 m, was geometrically corrected, using nearest-neighbor resampling, to a 3-metre-resolution raster image consisting of a mosaic of several 1:10,000 scale Ontario Base Maps. Although the OBM maps had an original resolution of 1 m, a 3-m resolution was chosen for the raster image as a compromise between geometric precision and file size. The geometrically-corrected radar image was then resampled to a resolution of 6.25 metres (the resolution of the raw data) using cubic convolution.

The resulting imagery was spatially filtered using a 5-by-5-pixel median filter to reduce image speckle and lower the standard deviation of crop signatures (Hutton and Brown, 1986).

Statistics Generation

On imagery from each of the five dates, training areas for corn, soya bean, tobacco, and winter wheat crops were created over fields where crop conditions such as growth stage and percent ground cover were quite uniform. (See Table 1 for a summary of all crop conditions considered.) The mean and standard deviation of the median-filtered, narrow-mode image data were generated and plotted (Table 2 and Figure 2).

RESULTS AND DISCUSSION

The Effect of Crop Cover Type

Winter wheat was the only crop which could be accurately identified as a class separate from other crop classes. The mean value of the radar return from winter wheat was found to be significantly greater than that from corn, soya beans and tobacco for at least one of the radar bands on four of the five dates. There is little difference between the winter wheat signature and the other crop signatures on the July 18 date, when all of the crops had matured enough to cover much of the ground on which they were growing, and the winter wheat had not yet been harvested. After the winter wheat was harvested, its mean radar return was much lower than that of the other crops.

The mean values of corn, soya bean and tobacco crops for all bands are different, but not significantly so because of the wide standard deviation of the signatures. The wide standard deviation of crop signatures is characteristic of radar data. In addition, a drought in the study area in 1988 caused crop growth to vary by soil type and topography more than in other years, thus increasing the standard deviation of the signatures.

The greatest difference in mean radar return values between the corn, soya bean and winter wheat crops was found on C-band HH and VV-polarized data recorded on August 3, the latest of the five image dates. By this time, the corn, soya beans and tobacco were approaching peak growth, and the winter wheat had been harvested. The greatest mean value for corn among all the data sets was obtained on the C-HH data from this date, while the greatest mean value for soya beans occurred on the C-VV data. The differences in structure between corn and soya bean plants may account for the differences in return between the HH and VV polarized radar.

The greatest mean value for tobacco among all the data sets was obtained on the August 3 X-band VV-polarized data, but, the greatest difference between tobacco and the other crops was found on a combination of the two August 3 C-band data sets. The difference among tobacco, corn and winter wheat was greatest on the HH-polarized data, while the difference between tobacco and soya beans was greatest on the VV-polarized data. There was little difference between the C-band HH and VV-polarized radar signatures of tobacco.

The mean value for winter wheat was significantly lower than that of the other three crops on both the C-HH and C-VV data, because the winter wheat crop had been harvested by the August 3 image date, leaving only dry stubble on the fields.

It is expected that, for any crop, the standard deviation of field-by-field mean image data values will be much less than the standard deviation of image data within fields. If that is so, a classification of fields by their mean values might be used to classify the four crops in the study area with greater accuracy than can be achieved using a standard pixel-by-pixel classification.

The Effect of Date

The main effect of the date of radar image acquisition lies in the percentage of ground surface covered by the crops. Ulaby *et al.* (1986) found that backscatter values increased as corn and winter wheat crops covered a greater percentage of ground, up to the point when the crops reached maturation. After that time, their backscatter values decreased. Similar results were found in this study. The backscatter values for corn, soya bean and tobacco crops increased steadily from the first to the fifth image recorded. As these crops had not yet reached the peak in their growth stages by the fifth image date, no decrease in backscatter was detected. The backscatter values for winter wheat fields increased from the first to the fourth image date, but decreased for the fifth image date, when the wheat crop had been harvested.

The Effect of Radar Wavelength

The greatest effect of wavelength was found in winter wheat signatures on the June 7 imagery. The mean C-band radar return values were significantly greater than the mean X-band values. The mean value for winter wheat on the C-band data was also significantly greater than the value for the other three crops; on the X-band data, however, there was no significant difference among the mean values for the four crops.

The longer C-band waves would be expected to penetrate more of the crop canopy and to be reflected by more of the crop structure below the canopy than the shorter X-band waves, which would be expected to reflect mainly off the surface of the canopy (NASA/JPL, 1986). Winter wheat may have had higher C-band values because the crop canopy reflected fewer radar waves than the underlying crop structure.

The Effect of Polarization

It was found that, generally, with the exception of the August 3 imagery of corn crops, the mean return values of vertically polarized radar were slightly greater than the mean return values of horizontally polarized radar. The vertically polarized C-band and X-band image data of corn has a lower mean value than that of the horizontally polarized image data.

NASA/JPL (1986) and Ulaby, *et al.*, (1986) found that the penetration of a vertically-polarized radar signal directed down through the structure of vertically-oriented crops such as winter wheat was less than that of a horizontally-polarized signal. Assuming that a lower measurement of signal penetration under such circumstances represents a higher signal reflectance, the greater return of vertically-polarized radar waves from crops other than corn found in the present study may be caused by the mainly vertical orientation of those crops. Perhaps the leaf portion of the corn crop, which is

more horizontal than vertical, affects the return of radar waves more than the vertically-oriented stalk portion. This difference in surface geometry between corn and the other crops would explain why the HH return from corn is greater than the VV return.

SUMMARY AND CONCLUSIONS

Although the set of radar images that was analysed was incomplete, and the effects of crop orientation and slope aspect were not analysed, the following conclusions may still be drawn from this preliminary analysis of the available data:

Significant differences were found between the radar image signatures of winter wheat and those of corn, soya beans and tobacco for all image dates other than July 18. The greatest differences in radar image signatures between corn, soya beans, and tobacco were found in C-band HH and VV-polarized data recorded on August 3, when these crops were at or near peak growth.

The radar backscatter from all crops increases as the crops approach maturity later in the growing season. The radar backscatter from harvested winter wheat is significantly lower than that from winter wheat approaching maturity.

The greatest effect of radar wavelength was found in winter wheat signatures taken from both HH and VV-polarized imagery, recorded on June 7. The mean C-band value from winter wheat was significantly greater than that of the other three crops, while there was no significant difference among the X-band mean values from all four crops.

The mean return values of vertically-polarized radar were generally slightly higher than the mean return values of horizontally-polarized radar, with the exception of the corn-crop value recorded on August 3, when the mainly horizontal leaf layer of the corn crop would have been sufficiently thick to modify the return.

Ongoing studies will include analysis of the complete radar data set and the consideration of other factors such as crop orientation and slope aspect. Field-by-field classification of the imagery on a geographic information system will be evaluated, and the degrees of correlation between radar reflectance and the factors that affect it will be determined.

ACKNOWLEDGEMENTS

The Canada Centre for Remote Sensing acquired the radar data used in this study. The airborne sensing unit of OCRS acquired the supporting aerial photography. The Ontario Ministry of Agriculture and Food, the University of Guelph and the University of Waterloo provided valuable assistance in field survey.

REFERENCES

Hutton, C. A. and R. J. Brown, "Radar backscatter dependence upon look direction", Proceedings of the Eleventh Canadian Symposium on Remote Sensing, Vol. 2, pp.541-558, 1987.

NASA/JPL, Shuttle Imaging Radar - C Science Plan, JPL Publication 86-29, Jet Propulsion Laboratory, Pasadena, CA., 1986.

Ulaby, F. T., R. K. Moore and A. K. Fung, "From Theory to Applications", Microwave Remote Sensing - Active and Passive, Vol.3, pp.1065-2162, 1986.

Wicklund, R. E. and N. R. Richards, The Soil Survey of Oxford County, Report No. 28 of the Ontario Soil Survey, Research Branch, Canada Department of Agriculture and the Ontario Agriculture, 1961.

Figure 1 STUDY AREA LOCATION MAP

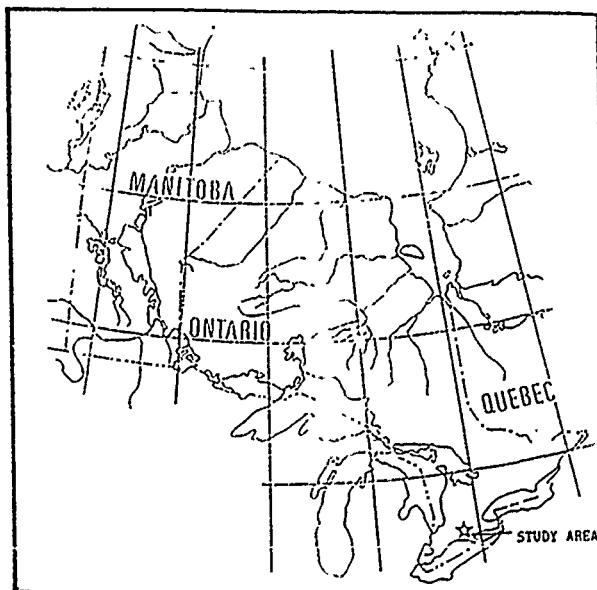


Table 1

STUDY AREA FIELD CONDITIONS

| DATE | CORN | | SOYA BEANS | | GROWTH STAGE |
|----------|---------------|--------|---------------|--------|--------------|
| | CROP HT. (cm) | COVER% | CROP HT. (cm) | COVER% | |
| 88-05-04 | 0 | 0 | 0 | 0 | bare soil |
| 88-06-02 | 15 | 0-5 | 0-5 | 0-2 | emergence |
| 88-07-18 | 100-140 | 70-80 | 30-50 | 75-85 | vegetative |
| 88-08-03 | 190-220 | 100 | 60-80 | 80-100 | seed devel. |
| DATE | TOBACCO | | WINTER WHEAT | | GROWTH STAGE |
| | CROP HT. (cm) | COVER% | CROP HT. (cm) | COVER% | |
| 88-05-04 | 0 | 0 | 10 | 20 | vegetative |
| 88-06-02 | 10 | 0 | 55-65 | 90 | vegetative |
| 88-07-18 | 85-95 | 85-90 | 80 | 95 | senescing |
| 88-08-03 | 140-160 | 90-100 | 20 | 0 | harvested |

Table 2

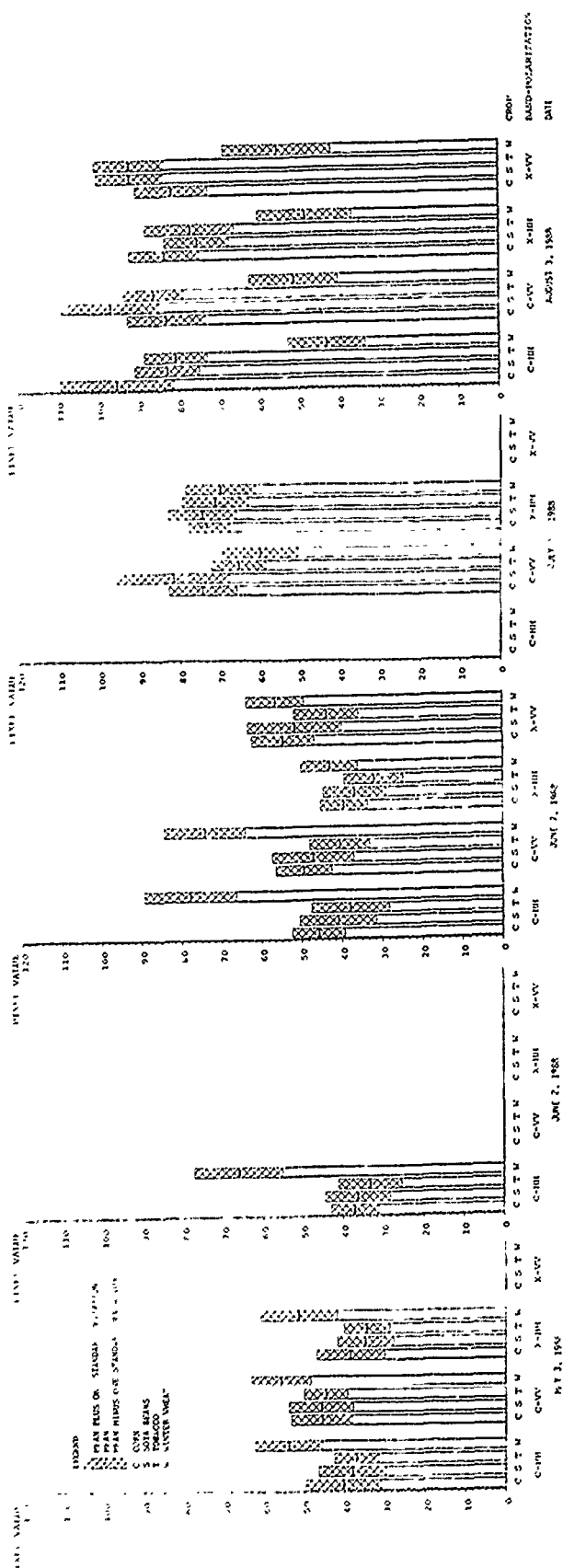
RADAR IMAGE CROP SIGNATURES

| DATE | CROP | C-HH | | C-VV | | X-HH | | X-VV | |
|----------|--------------|------|---------|------|---------|------|---------|------|---------|
| | | MEAN | ST. DV. | MEAN | ST. DV. | MEAN | ST. DV. | MEAN | ST. DV. |
| 88-05-03 | CORN | 40.2 | 9.4 | 45.6 | 7.6 | 38.5 | 8.4 | . | . |
| | SOYA BEANS | 38.0 | 8.4 | 45.9 | 8.0 | 34.8 | 7.0 | . | . |
| | TOBACCO | 37.0 | 5.5 | 44.7 | 5.4 | 34.5 | 5.7 | . | . |
| 88-06-02 | WINTER WHEAT | 53.9 | 8.5 | 55.8 | 7.4 | 51.5 | 9.6 | . | . |
| | CORN | 37.1 | 5.9 | . | . | . | . | . | . |
| | SOYA BEANS | 36.4 | 8.1 | . | . | . | . | . | . |
| 88-06-07 | TOBACCO | 33.4 | 7.9 | . | . | . | . | . | . |
| | WINTER WHEAT | 66.3 | 11.0 | . | . | . | . | . | . |
| | CORN | 45.9 | 6.5 | 49.5 | 6.9 | 39.5 | 5.8 | 54.8 | 7.8 |
| 88-07-18 | SOYA BEANS | 41.1 | 9.4 | 47.2 | 10.1 | 36.9 | 7.8 | 51.7 | 11.8 |
| | TOBACCO | 38.0 | 9.6 | 40.6 | 7.4 | 32.2 | 7.3 | 43.8 | 8.0 |
| | WINTER WHEAT | 77.9 | 11.5 | 74.2 | 10.1 | 43.2 | 7.1 | 56.5 | 7.2 |
| 88-08-03 | CORN | . | . | 74.3 | 8.4 | 73.5 | 5.8 | . | . |
| | SOYA BEANS | . | . | 81.7 | 14.1 | 74.6 | 8.5 | . | . |
| | TOBACCO | . | . | 65.4 | 6.9 | 71.4 | 8.3 | . | . |
| 88-08-03 | WINTER WHEAT | . | . | 59.9 | 9.5 | 70.1 | 8.8 | . | . |
| | CORN | 95.4 | 14.1 | 83.0 | 9.7 | 83.6 | 8.7 | 81.7 | 9.0 |
| | SOYA BEANS | 82.9 | 8.0 | 97.0 | 12.2 | 75.6 | 8.0 | 92.1 | 8.4 |
| 88-08-03 | TOBACCO | 80.8 | 7.8 | 86.3 | 7.4 | 77.4 | 11.0 | 92.2 | 8.8 |
| | WINTER WHEAT | 43.2 | 9.6 | 51.5 | 11.0 | 48.7 | 11.8 | 55.5 | 13.4 |

Figure 2

GRAPHS OF RADAR IMAGE MEAN AND STANDARD DEVIATION VALUES

CROP TYPE, BAND-POLARIZATION AND DATE



Recognition of three-dimensional objects
from single images

David G. Lowe
Assistant Professor and CIAR Scholar
Department of Computer Science
University of British Columbia

Recent work in the area of computer-based visual recognition shows that it is possible to produce reliable identification and localization of known objects from even partial image data. In this talk, a computer vision system will be described that can recognize three-dimensional modeled objects from unknown viewpoints in ordinary gray-scale images. This system, named SCERPO, is quite robust in the sense that it is insensitive to partial occlusion and noisy data. An important requirement for achieving robustness is the application of the viewpoint consistency constraint. This constraint requires that the locations of all object model features in an image must be consistent with projection of the model from a single viewpoint. Unfortunately, while simple to state, this constraint presents a number of mathematical difficulties when reasoning between 3-D models and 2-D images. We will describe how the constraint can be implemented and also used to greatly reduce the search space during model-based matching. A second component of the system is based on the principle of detecting structure in the image that is present over a range of viewpoints of the known objects and yet is unlikely to have arisen by accident. Specific methods have been developed for measuring the significance of instances of collinearity, proximity of endpoints, and parallelism between straight line segments. Using these results, we have developed algorithms that can identify these types of groupings in an image and then match these derived image structures to corresponding structures of a model. Since these image groupings reflect viewpoint-invariant aspects of a three-dimensional scene, they are ideal structures for bridging the gap between the two-dimensional image and the three-dimensional model. By combining the use of viewpoint independent groupings to index into a set of object descriptions with the precise determination of viewpoint for confirmation, it is possible to create a robust system for visual recognition.

Frame Camera and Image Contextual Constraints for Stereo Matching in Aerial Scenes

DAVID W. PAGLIERONI*

Ford Aerospace Corporation
220 Henry Ford II Drive
San Jose, CA 95134

Abstract Many remote sensing and reconnaissance applications, including image ortho-rectification and registration, geopositioning and mensuration, utilize geodetic coordinate databases. Such 3-D information is often obtained manually or semi-automatically from stereo pairs of optical aerial images. Extracting accurate geodetic data for a given point requires its *conjugate point*, the camera model and precise knowledge of all camera parameters. Using nominal parameter values and *ground control* or *ground truth* for at least 3 imaged points, precise parameter values can be computed by an iterative process known as *resection*. Automatic geodetic data extraction can thus be approached in terms of *stereo matching*, i.e. by determining matches in one image to arbitrary pixels in the other. This paper addresses the specific problem of pixel-by-pixel stereo matching in digitized aerial scenes acquired by frame cameras.

First, a new stereo feature matching algorithm is developed. Using distance transforms and analytical optimization procedures, it adaptively matches pairs of feature sets and then, unlike other algorithms, applies the camera model to validate the results. Its advantages are efficiency, amenability to parallelism and match consistency with respect to both image context and camera parameters. Furthermore, the algorithm uses only one edge resolution and despite absence of structural feature descriptions, is not troubled by special feature configurations.

Second, the feature matches are used by a novel procedure for rapid computation of matches to *arbitrary* pixels. One matched point coordinate is estimated as a weighted sum of surrounding feature match coordinates and then the other is specified by the camera constraint. The resulting surfaces are efficiently generated, usually smooth, based on matched pixels with lines-of-sight that intersect and basically have the correct geodetic coordinates at all previously matched feature pixels.

Keywords. stereo matching frame camera features correlation predictor

1. Introduction

Many remote sensing and reconnaissance applications, including image ortho-rectification and registration, geopositioning and mensuration, utilize geodetic coordinate databases. Such 3-D information is often obtained manually or semi-automatically from stereo pairs of optical aerial images. Extracting accurate geodetic data for a given point requires its *conjugate point*, the camera model and precise knowledge of all camera parameters. Using nominal parameter values and *ground control* or *ground truth* at $n \geq 3$ imaged points, precise parameter values can be computed by an iterative process known as *resection* [1]. Automatic geodetic data extraction can thus be approached in terms of *stereo matching*, i.e. by determining matches in one image to arbitrary pixels in the other. This paper addresses the specific problem of pixel-by-pixel stereo matching in digitized aerial scenes acquired by frame cameras.

One approach to stereo matching, though computationally intensive, involves cross-correlating stereo pair patches under presumption of pair geometrical consistency in all but offset, adequate feature content within blocks and sufficient gray-scale similarity between block pairs. Another approach to 3-D data extraction attempts to compute reliable altitudes for feature pixels and then interpolates in-between. This is consistent with the fact that stereo matches to arbitrary pixels, even when obtained manually, are often just estimates. However, existing *stereo feature matching* algorithms for computing feature altitudes are computationally intensive and classical schemes for interpolating nonuniformly scattered variable altitude posts, despite resulting in surfaces which miss the posts, are only suited to massively parallel implementation. The stereo matching system presented here is more consistent with the latter approach in that stereo matches to arbitrary pixels are based partially on surrounding feature matches, but they are also forced to satisfy an independent camera constraint.

First, a new stereo feature matching algorithm is developed. Using distance transforms and analytical optimization procedures, it adaptively matches pairs of feature sets and then, unlike other algorithms, applies the camera model to validate the results. Its advantages are efficiency, amenability to parallelism and match consistency with respect to both image context and camera parameters. Furthermore, the algorithm uses only one edge resolution and despite absence of structural feature descriptions, is not troubled by special feature configurations.

Second, the feature matches are used by a novel procedure for rapid computation of matches to *arbitrary* pixels. One matched point coordinate is estimated as a weighted sum of surrounding feature match coordinates and then the other is specified by the camera constraint. The resulting surfaces are efficiently generated, usually smooth, based on matched pixels with lines-of-sight that intersect and basically have the correct geodetic coordinates at all previously matched feature pixels.

2. Stereo Feature Matching

Features, such as edge pixels, are easier to match than other points because they are salient and relatively consistent between stereo images acquired in succession. Stereo feature matching has thus become an active research topic. The objective is to determine reliable feature rather than exhaustive pixel correspondences and matches to feature pixels in one stereo image may or may not be limited to feature pixels in the other. *Motion stereo methods* [2-3] match feature in overlapping scenes by applying simplified matching procedures to all successive pairs of intermediate frames, but require many images. Techniques that apply in the absence of intermediate frames are often computationally intensive. One applies the Marr-Poggio computational human stereo vision model to establish consistent unambiguous stereo edge associations, but it is troubled by certain feature configurations such as horizontal lines [4]. Others

* This research was funded under Independent Research and Development by Ford Aerospace Corporation

match higher level feature descriptions, where structural description is an on-going research topic [5]. A new adaptive algorithm that applies distance transforms and analytical optimization procedures to the problem of matching edge sets is presented in this section.

2.1 Frame Camera Model

Before attempting to match features, it is important to correct geometrical distortions between stereo pairs. These distortions are normally dominated by offset and possibly aspect ratio variation along the parallax direction. However manifested, the camera model can be used to reduce them.

Frame camera models are specified by *exterior orientation parameters*, which define camera position and pointing direction relative to some reference world coordinate frame at the instant of image acquisition, and *interior orientation parameters*, which define the transform between pixels and camera frame coordinates [1]. The 6 exterior orientation parameters are $(x_0, y_0, z_0) \equiv$ camera xyz position in world coordinates (meters) and $(\phi, \omega, \kappa) \equiv$ camera orientation angles relative to the world frame (rad) where the camera to world frame rotation matrix $R(\phi, \omega, \kappa)$ has the property that $R^{-1}(\phi, \omega, \kappa) = R(-\phi, -\omega, -\kappa) = R^T(\phi, \omega, \kappa)$.

Interior orientation can be handled by setting up an affine transform between camera frame coordinates (u, v, f) and column-row pixels (c, r) where u and v are focal plane coordinates (meters) and f is the camera focal length (meters), is an interior orientation parameter. In this case, the lists $[c, r] \equiv \{c(i), r(i) \mid i=0, \dots, n-1\} \equiv n_0 \geq 3$ column-row pixel coordinates and $[u, v] \equiv \{u(i), v(i) \mid i=0, \dots, n-1\} \equiv$ corresponding focal plane coordinates (meters) are used in computing least-squares w and s coefficients where

$$u = w(0,0)c + w(1,0)r + w(2,0) \quad (2.1a)$$

$$v = w(0,1)c + w(1,1)r + w(2,1) \quad (2.1b)$$

$$c = s(0,0)u + s(1,0)v + s(2,0) \quad (2.2a)$$

$$r = s(0,1)u + s(1,1)v + s(2,1) \quad (2.2b)$$

These coefficients can be viewed as the remaining interior orientation parameters

Let $I \equiv$ camera index (L for left, R for right), $P_I \equiv [x_I, y_I, z_I]^T \equiv$ camera I world frame position (meters), $R_I \equiv$ camera I -to-world frame rotation matrix, $f_I \equiv$ camera I focal length (meters) and $(u_I, v_I) \equiv$ focal plane coordinates (meters) for point imaged by camera I . For frame cameras, the 3-D lines in the world frame, through focal plane points (u_I, v_I) are

$$l_i = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mid \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix} + t_1 R_I \begin{bmatrix} u_I \\ v_I \\ f_I \end{bmatrix} \mid t_1 \in \mathbb{R} \right\} \quad I = L \text{ or } R \quad (2.3)$$

where t_1 is the parameter along l_i . Now choose $n \geq 3$ terrain points with well distributed UTM northings and eastings known to lie in the area of interest and with fixed assumed altitudes. Let $P(i) \equiv [x(i), y(i), z(i)]^T \mid i=0, \dots, n-1$ be their world (i.e. *USR* or *geocentric*) coordinates. Then, with $R_I \equiv [R_{I,0}, R_{I,1}, R_{I,2}]$, the associated focal plane coordinates $(u_I(i), v_I(i))$ are given by

$$u_I(i) = u_I(P(i)) = \frac{R_{I,0}^T [P(i) - P_I] f_I}{R_{I,2}^T [P(i) - P_I]} \quad (2.4a)$$

$$v_I(i) = v_I(P(i)) = \frac{R_{I,1}^T [P(i) - P_I] f_I}{R_{I,2}^T [P(i) - P_I]} \quad (2.4b)$$

(2.2) can then be used to generate $\{c_i(t), r_i(t) \mid i=0, \dots, n-1\}$ for $I=L$ and R so that as in (2.2), least-squares transform coefficients between these two sets of pixels can be derived.

2.2 Adaptive Feature Matching Algorithm

Fig 1(c-d) are the $\sigma=3$ pixel LoG edge maps based on the strongest 40% zero crossings that lie on segments longer than 35 pixels for the high terrain relief stereo pair of Fig. 1(a-b) acquired by frame camera. The right image was warped using the 6 coefficients derived in section 2.1 and the 512×512 left image was split into 16 contiguous 128×128 blocks. Each block's vertical offset relative to the warped right image was computed via epipolar line formula (section 2.3) while horizontal offset was approximated via normalized cross correlation along the epipolar line. Left image features were offset

accordingly prior to feature matching.

The nearest horizontal feature transform $N(c, r) \equiv [n_c(c, r), n_r(c, r)] \equiv$ feature pixel on row r closest to (c, r) was then computed for the warped right feature map. If a given row has no feature pixels, n_c does not exist. For each 128×128 block, let P be the list of offset left image feature pixels for which n_c exists and let N ($m \times 2$) be the nearest horizontal feature transform of P in the warped right image. Initially, $P = P_0$ and $N = N_0$. Let $f(P, t)$ ($m \times 2$) be the result of spatially transforming P in accordance with adjustment parameters t . With N_k defined as the nearest horizontal feature transform of P_k , $P_{k+1} = f(P_k, t_{k+1})$ and $\langle a_x, a_y \rangle \equiv [\|a_x\|^2 + \|a_y\|^2]^2$ (a_x, a_y , $m \times 1$), consider an algorithm where at iteration $k+1$, the problem is to analytically compute t_{k+1} minimizing $\langle f(P_k, t_{k+1}) - N_k \rangle$. This iterative process is adaptive in that if P_{k+1} has a different nearest feature transform than P_k , the algorithm causes P_{k+1} to adapt to N_{k+1} on the next iteration. However, adaptation seeks out local match minima so $(s+1) > 1$ sets of initial parameters are required.

Because terrain relief produces distortion only along the epipolar line, which is locally horizontal, t should be taken as the 1×2 vector of column offset and aspect ratio parameters, i.e., $t \equiv [T_c, A_c]$ and

$$f(P, t) = (P - [T_c, 0]) \begin{bmatrix} A_c & 0 \\ 0 & 1 \end{bmatrix} \quad (2.5a)$$

where T_c is the $m \times 1$ vector of elements T_c , the first column of P contains pixel column coordinates, i.e., $P = [c, r]$, and $N \equiv [n_c, n_r]$. Then the t minimizing $\langle f(P, t) - N \rangle$ is

$$t = [T_c, A_c] = [\bar{n}_c - \bar{c}, (\eta_c^T c - m \bar{\eta}_c \bar{c}) / (\|c\|^2 - m \bar{c}^2)] \quad (2.5b)$$

where \bar{n}_c is the mean n_c element and \bar{c} is the mean c element. Then for each block, the adaptive optimization algorithm is

for $i=0, \dots, s$

$$t_0(i) = [T_c(i), A_c(i)] = [(i - \frac{s}{2}) \Delta, 1] \text{ (say } \Delta = 5 \text{ pixels)}$$

$$P_0(i) = f(t_0(i), P)$$

while $(k=0 \text{ or } \langle P_k(i) - P_{k-1}(i) \rangle > \epsilon)$

compute $t_{k+1}(i)$ minimizing $\langle f(P_k(i), t_{k+1}(i)) - N_k(i) \rangle$

$P_{k+1}(i) = f(P_k(i), t_{k+1}(i))$, increment k

$t(i) = t_k(i)$, $P(i) = P_k(i)$, $d(i) = \langle P_k(i) - N_k(i) \rangle$

$i^* = i$ minimizing $d(i)$, $t^* = P(i^*) = f(P_0, t^*)$

This algorithm guarantees improvement or convergence at each iteration since $\langle P_{k+1} - N_k \rangle \leq \langle P_k - N_k \rangle$, because t_{k+1} is the t minimizing $\langle f(P_k, t) - N_k \rangle$ and $\langle P_{k+1} - N_{k+1} \rangle \leq \langle P_{k+1} - N_k \rangle$, by definition of nearest horizontal feature transform. Therefore, $\langle P_{k+1} - N_{k+1} \rangle \leq \langle P_k - N_k \rangle$ and convergence occurs at the first iteration k for which $\langle P_{k+1} - P_k \rangle < \epsilon$ (say $\epsilon = 1$ pixel).

Matches are checked for consistency by analyzing the distribution of horizontal disparities for each left feature segment. Multiple left feature matches to the same right feature are removed by choosing the closest one and, if too far away, the closest ones are also removed. A list of candidate matches composed of the remaining right feature pixels prior to geometrical correction and their left counterparts is then assembled.

2.3. Frame Camera Constraint Line

Bad candidate feature matches cannot always be eliminated using gray scale contextual information alone. In addition to being contextually consistent, valid feature matches must also satisfy an independent camera constraint. Through the notion of miss-distance, the frame camera model provides a convenient mechanism for validating candidate matches.

Two points cannot be conjugate if the miss-distance between their lines of sight is too large. Let

$$Q_i \equiv [\hat{u}_i, \hat{v}_i, \hat{f}_i]^T \equiv R_i [\hat{u}_i, \hat{v}_i, \hat{f}_i]^T \quad I = L \text{ or } R. \quad (2.6)$$

For frame cameras, the set of all right image focal plane points (u_R, v_R) producing 0 miss-distance lines with a given left image focal plane point (u_L, v_L) is a line in the right focal plane with slope and intercept that are functions of u_L and v_L . Specifically, it can be shown that the frame camera constraint (epipolar) line is

$$u_R = a(u_L, v_L)v_R + b(u_L, v_L) \quad (2.7a)$$

$$a(u_L, v_L) = \frac{(R_{R,1} \times Q_L)^T \Delta P}{(Q_L \times R_{R,0})^T \Delta P} \quad (2.7b)$$

$$b(u_L, v_L) = \frac{(R_{R,2} \times Q_L)^T \Delta P}{(Q_L \times R_{R,0})^T \Delta P} f_R \quad (2.7c)$$

where $R_R = [R_{R,0}, R_{R,1}, R_{R,2}]$. Furthermore,

$$t_L = \frac{(\Delta P \times R_{R,1})^T R_{R,0} v_R + (\Delta P \times R_{R,2})^T R_{R,0} f_R}{(Q_L \times R_{R,1})^T R_{R,0} v_R + (Q_L \times R_{R,2})^T R_{R,0} f_R} \quad (2.8a)$$

$$t_R = \frac{(Q_L \times R_{R,0})^T \Delta P}{(Q_L \times R_{R,1})^T R_{R,0} v_R + (Q_L \times R_{R,2})^T R_{R,0} f_R} \quad (2.8b)$$

A 0 miss-distance frame camera constraint can be imposed on feature matches (u_L, v_L) and (u_R, v_R) (obtained as in section 2.2) by computing the focal plane point (u_R, v_R) on the 0 miss-distance line associated with (u_L, v_L) closest to (u_R, v_R) . The solution is

$$v_R = \frac{a(u_L, v_L)[u_R - b(u_L, v_L)] + v_{R,0}}{a^2(u_L, v_L) + 1} \quad (2.9)$$

with u_R then given by (2.7). Candidate matches for which $<[c_{R,0}, r_{R,0}], [c_R, r_R]> > \Delta d$ (where the pixels are computed via (2.2b)) are eliminated. The stereo matches obtained via (2.9) have associated miss-distances of 0, as they must, and their world coordinates, which can readily be converted to UTM, are given by applying (2.8) to (2.3). The stereo matches produced by applying this algorithm to the images of Fig. 1(a-b) are shown in Fig. 1(e-f). Note how well feature structure is preserved even in the presence of high terrain relief.

3. Stereo Matching

The goal of stereo feature matching is to determine points conjugate to reliable feature pixels. However, many applications require 3-D information at regular intervals or even on a pixel-by-pixel basis. Thus, the ability to efficiently perform stereo matching at arbitrary pixels is important.

There are two basic approaches for extracting 3-D data at arbitrary pixels from stereo pairs. One computes matches to arbitrary pixels by cross-correlating blocks centered at those pixels with the other image along epipolar lines [6-7]. It fails in regions void of features and areas of high terrain relief containing shadows, occlusion or geometrical distortion. On a point-by-point basis, the cost per point is $O(b_1 b_2)$ adds and multiplies where b_1 is the block width and b_2 is the search width. However, this cost can be reduced to $O(b_1 b_2)$ adds and multiplies per point when all points along a given scan line are processed together.

The second approach matches features, computes their world coordinates and interpolates to estimate world coordinates at arbitrary pixels in between. Unlike correlation, this method produces reasonable world coordinate estimates for pixels in regions absent of features. Existing techniques (e.g. [8-10]) interpolate between nonuniformly distributed variable altitude posts by minimizing special functionals (perhaps with penalty terms). Though smooth, these surfaces often miss altitude posts and are only suited to massively parallel implementation.

The algorithm developed here uses surrounding feature matches to efficiently compute matches, with intersecting lines of sight, to arbitrary pixels. It produces results comparable to correlation in areas of high feature activity and deals with the other areas as well. The frame camera constraint equation is first re-written as

$$\mu_R = a(u_L, v_L)v_R + \beta(u_L, v_L) \quad (3.1a)$$

$$(\mu_R, v_R, a(u_L, v_L), \beta(u_L, v_L)) = \quad (3.1b)$$

$$\begin{cases} (u_R, v_R, a(u_L, v_L), \beta(u_L, v_L)) & |a(u_L, v_L)| \leq 1 \\ (v_R, u_R, \frac{1}{a(u_L, v_L)}, -\beta(u_L, v_L)) & |a(u_L, v_L)| > 1. \end{cases}$$

ν disparities for arbitrary points are estimated in terms of known ν disparities for neighboring points. μ is not used because $|a(u_L, v_L)| \leq 1 \Rightarrow$ estimation errors in μ_R can, via (3.1a), produce significantly larger errors in v_R . Matches to arbitrary pixels are forced to have intersecting lines of sight by applying the frame camera constraint (3.1a) to calculation of μ_R . The advantage of this approach over altitude interpolation is that using (2.8) and (2.3), all

3 world coordinates can be readily derived and made consistent with intersecting lines of sight.

Suppose that it is desired to estimate the ν coordinate, $\nu_R(c_L, r_L)$, of the stereo match to arbitrary left image pixel (c_L, r_L) . With $u_L(c_L, r_L)$ and $v_L(c_L, r_L)$ as in (2.2), $\delta(c_L, r_L)$ defined as the ν disparity associated with (u_L, v_L) and (μ_L, ν_L) defined consistent with (μ_R, ν_R) , we have

$$\nu_R(c_L, r_L) = \nu_L(c_L, r_L) + \delta(c_L, r_L) \quad (3.2)$$

so that only δ need be estimated. With

$$(c_L)_{\min} = \max(c_{\min}, c_L - b), \quad (c_L)_{\max} = \min(c_{\max}, c_L + b) \quad (3.3a)$$

$$(r_L)_{\min} = \max(r_{\min}, r_L - b), \quad (r_L)_{\max} = \min(r_{\max}, r_L + b) \quad (3.3b)$$

let us estimate $\delta(c_L, r_L)$ for image pixels $(c_L, r_L) \in [c_{\min}, c_{\max}] \times [r_{\min}, r_{\max}]$ in terms of δ values for certain pixels in $[c_L, c_{\min}], [c_L, c_{\max}] \times [r_L, r_{\min}], [r_L, r_{\max}]$ at which δ is known by stereo feature matching or has been previously estimated. With $R(c_L, r_L)$ (discussed later in this section) defined as this set of pixels and

$$D[(c_L, r_L), (c, r)] = \text{absolute distance from } (c_L, r_L) \text{ to } (c, r) \quad (3.4a)$$

$$D_{\min}(c_L, r_L) = \min_{(c, r) \in R(c_L, r_L)} D[(c_L, r_L), (c, r)] \quad (3.4b)$$

consider the class of δ estimates

$$\hat{\delta}(c_L, r_L) = \sum_{(c, r) \in R(c_L, r_L)} w(c, r, c_L, r_L) \delta(c, r) \quad (3.5a)$$

$$w(c, r, c_L, r_L) = \frac{D_{\min}(c_L, r_L) + \epsilon}{f D[(c_L, r_L), (c, r)] + \epsilon} w_0(c, r, c_L, r_L) \quad (3.5b)$$

where $f(0)=0$, f is an increasing function, $\epsilon > 0$ is some constant and $w_0(c, r, c_L, r_L)$ is chosen such that

$$\sum_{(c, r) \in R(c_L, r_L)} w(c, r, c_L, r_L) = 1 \Rightarrow \quad (3.5c)$$

$$w_0(c, r, c_L, r_L) = \frac{1}{\sum_{(c, r) \in R(c_L, r_L)} \frac{D_{\min}(c_L, r_L) + \epsilon}{f D[(c_L, r_L), (c, r)] + \epsilon}} \quad (3.5d)$$

Note that for $(c, r) \in R(c_L, r_L)$, $\delta(c, r)$ is already known, by stereo feature matching or prior estimation, and if $\epsilon \approx 0$, $\delta(c, r) \approx \delta(c, r)$, i.e., the estimator returns the correct geodetic coordinates at matched features. As ϵ increases, the surfaces smooth out more but computed coordinates at matched features tend to be somewhat off. A meaningful upper bound on ϵ can be derived by insisting that ϵ be small enough to ensure that weights attributable to all points in $R(c_L, r_L)$ except (c_L, r_L) sum to no more than a fraction p of the weight attributable to (c_L, r_L) when (c_L, r_L) already has a matched feature. The relation

$$\epsilon < \epsilon_{\max}(p, b) \equiv p/(2b+1)^2 \quad (3.6)$$

more than satisfies this requirement. Also, note that f needs to be an increasing function so that weights attributed to disparities associated with matched features decrease with increasing distance to the pixel of interest. Rapidly increasing functions promote high surface variation near matched features while slowly increasing functions enhance surface uniformity.

If b is not allowed to vary with (c_L, r_L) , some of the blocks may not contain matched feature pixels. While choosing b to be large reduces the likelihood of having empty blocks, it does not eliminate the problem, it increases cost and is inconsistent with the idea that disparity at (c_L, r_L) should be dictated only by the disparities of its nearest neighbors if they are known. On the other hand, allowing b to vary hampers implementability and can run surface integrity. By modifying concepts developed for 2-D linear prediction, it is possible to remove the empty block problem and still choose a small fixed value for b .

For this algorithm, consider 2 types of scans. For top-to-bottom and left-to-right scans from $(c_L, r_L) \in [1, c_{\max}-1] \times [1, r_{\max}-1]$, $R(c_L, r_L)$ will contain the 4 points (c_L-1, r_L) , (c_L+1, r_L-1) $i = -1, 0, 1$ in the causal prediction window of Fig. 2(a) [11] for which δ has been previously calculated during the scan together with matched feature pixels in $[c_L, (c_L)_{\max}] \times [r_L, (r_L)_{\max}] \cup [(c_L)_{\min}, c_L-1] \times [1, (r_L)_{\max}]$. For bottom-to-top and right-to-left scans, $R(c_L, r_L)$ will contain the 4 points (c_L+1, r_L) , (c_L-1, r_L+1) $i = -1, 0, 1$ in the causal prediction window of Fig. 2(b) for which δ has been previously calculated during the scan together with matched feature pixels in $[(c_L)_{\min}, c_L] \times [r_L, (r_L)_{\max}] \cup [c_L+1, (c_L)_{\max}] \times [(r_L)_{\min}, r_L-1]$.

Disparities are obtained by averaging results of 2 independent passes. With $(c_0, r_0) \equiv$ the leftmost matched feature pixel on the top row containing matches, the first pass computes row i , by left to

right and right to left scans from (c_0, r_0) . A top-to-bottom left-to-right scan from $(1, r_0+1)$ ($r_0 < r_{\max}-2$) and a bottom-to-top right-to-left scan from $(c_{\max}-2, r_0-1)$ ($r_0 > 1$) are then performed. Similarly, with $(c_i, r_i) \Leftarrow$ the rightmost matched feature pixel on the bottom row containing matches, the second pass completes row r_i by right-to-left and left-to-right scans from (c_i, r_i) . A bottom-to-top right-to-left scan from $(c_{\max}-2, r_i-1)$ ($r_i > 1$) and a top-to-bottom left-to-right scan from $(1, r_i+1)$ ($r_i < r_{\max}-2$) are then performed. The result is a relatively simple efficient scheme for sequential computation of high quality surfaces with correct disparities at the feature posts.

After stereo feature matching, the cost associated with computing matches to arbitrary pixels can be estimated in terms of the average number of feature pixels per $2b \times b$ block, which, for $N \times N$ images containing E feature pixels, is approximately $2Eb^2/N^2$. For the example of Fig. 1, $E=10520$, $N=512$ and $b=16$ so that the average number of feature pixels per $2b \times b$ block is $O(20)$. Using table look-up on f , the number of adds and multiplies required per disparity calculation is about 3 times this number or $O(60)$. Depending on b_1 and b_2 , the correlation approach can be 1 to 3 orders of magnitude more expensive. For example, $b_1=b_2=32 \Rightarrow O(32^2=1024)$ required operations per match which is 20 times more expensive for our example.

Results of applying (3.5) are shown in Fig. 1(g) where $f(x) = \sqrt{x}$ and $\epsilon = 1$. The surface altitude range is 1656 to 1788 meters. $n=32$ well distributed pixels, $\{(c_i(i), r_i(i)) \mid i=0, \dots, n-1\}$, were manually chosen, their manually selected counterparts, $(c_{RM}(i), r_{RM}(i))$, were compared to the automatically computed ones, $(\hat{c}_R(i), \hat{r}_R(i))$, using the pixel accuracy measure

$$Q_p = \left[\frac{1}{n} \sum_{i=0}^{n-1} \{(\hat{c}_R(i) - c_{RM}(i))^2 + (\hat{r}_R(i) - r_{RM}(i))^2\} \right]^{\frac{1}{2}} \quad (3.7)$$

and Q_p was found to be $O(4)$ pixels.

4. Conclusion

The camera model and its parameters contain much information and should be used to supplement gray-scale contextual data when computing conjugate pixels for stereo pairs. In stereo feature matching, bad candidate matches obtained on the basis of contextual information can be eliminated using the camera model and, when matching arbitrary pixels, it can be used to enforce line-of-sight consistency between conjugate points. The need for massively parallel networks in real-time computation of matches to large numbers of pixels can be eliminated by estimating, for each pixel, the focal plane coordinate dictated by the camera model as a weighted function of neighborhood data.

Acknowledgements

I would like thank Walt Eppler for the time he spent with me discussing this work, Nick Erndt for developing the graphics software and Sharon Tan for her help in implementing certain portions of the system.

References

- [1] Manual of Photogrammetry, Fourth Edition, American Society of Photogrammetry, pp.37-102, 1980.
- [2] R. Nevatia, "Depth Measurement by Motion Stereo", Comput. Graphics Image Processing, 6, pp.619-630, 1976.
- [3] Gang Xu, Tsuji Saburo and Minoru Asada, "A Motion Stereo Method Based on Coarse-to-Fine Control Strategy", IEEE Trans. PAMI, 9, 2, pp.332-336, March 1987.
- [4] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm", IEEE Trans. PAMI, 7, 1, pp.17-34, Jan. 1985.
- [5] K. L. Boyer and A. C. Kak, "Structural Stereopsis for 3-D Vision", IEEE Trans. PAMI, 10, 2, pp.144-166, March 1988.
- [6] M. P. Levine, D. A. O'Handley and G. M. Yagi, "Computer Determination of Depth Maps", Comput. Graphics Image Processing, 2, 2, pp.131-150, 1973.
- [7] U. V. Helava and W. E. Chapelle, Epipolar-Scan Correlation, Bendix Technical Journal, 5, 1, pp.19-23, 1972.
- [8] L. L. Schumaker, "Fitting Surfaces to Scattered Data", Approximation Theory II, G. G. Lorentz, C. K. Chui and L. L. Schumaker eds., Academic, 1976, pp.203-267.
- [9] W. E. L. Grimson, "An Implementation of a Computational Theory of Visual Surface Interpolation", CVGIP, 22, 1983, pp.39-69.
- [10] Demetri Terzopoulos, "The Computation of Visible Surface Representations", IEEE Trans. PAMI, 10, 4, July 1988, pp.417-438.
- [11] Anil K Jain, Fundamentals of Digital Image Processing, Prentice-Hall, 1989, p.204-219.

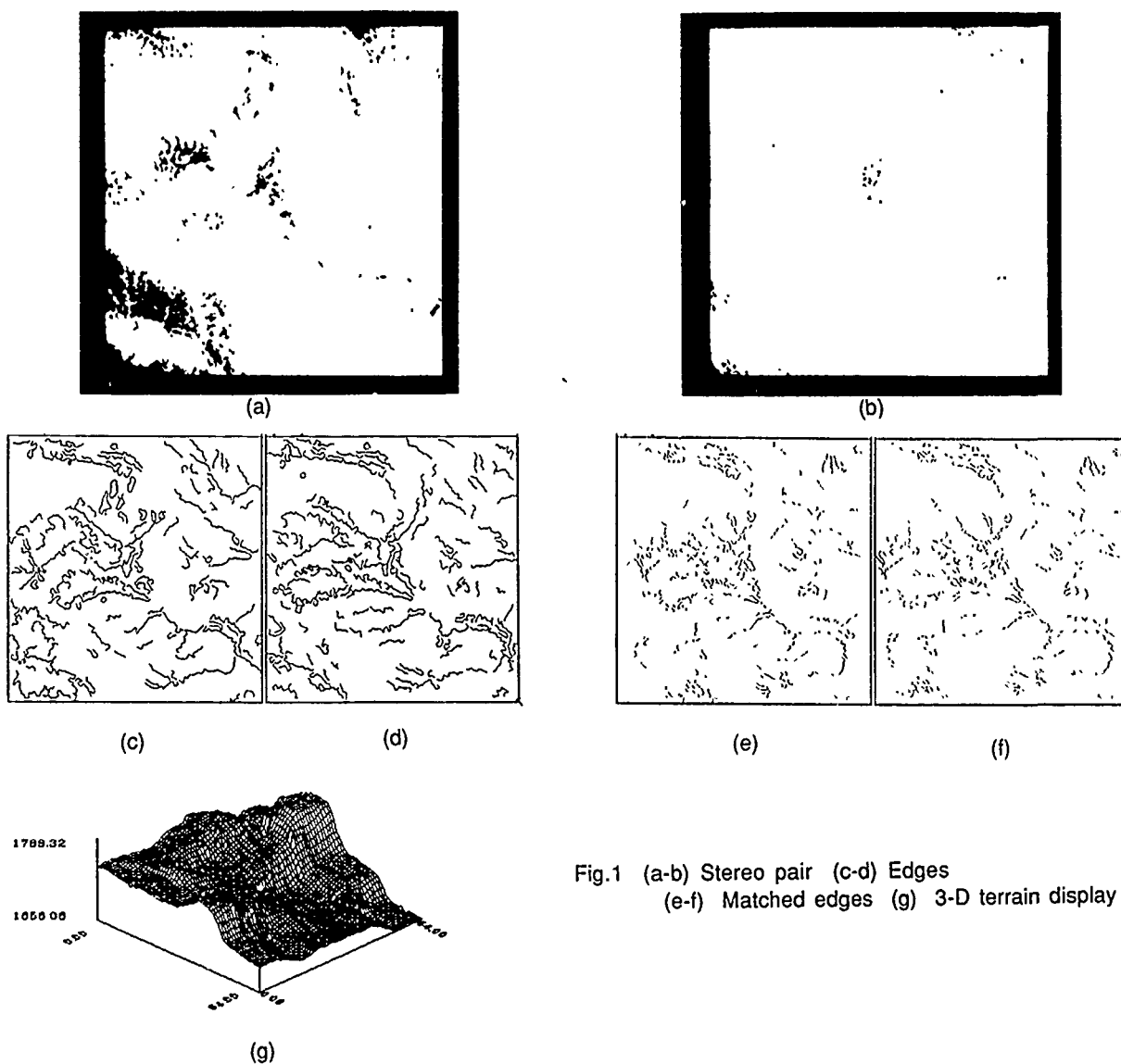


Fig.1 (a-b) Stereo pair (c-d) Edges
 (e-f) Matched edges (g) 3-D terrain display

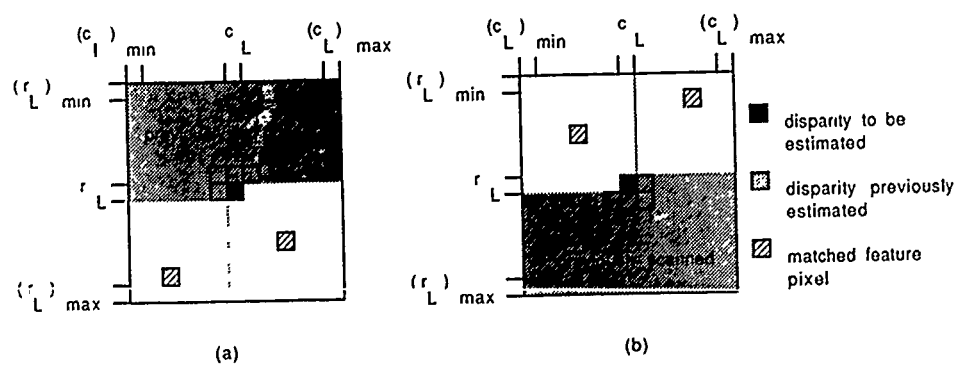


Fig.2 $R(c_L, r_L)$ geometry for row-wise (a) top-to-bottom left-to-right and
 (b) bottom-to-top right-to-left scans.

GEOMETRIC TRANSFORMATION AND TERRAIN CORRECTION OF REMOTELY-SENSED IMAGERY

J. Thomas Parr
The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, Massachusetts 01867
(617)-942-2000

ABSTRACT

The geometric adjustment or correction of remotely-sensed data is used in image processing for a variety of reasons. Most common is the desire to co-register multiple data sets, either from different sensors or from different dates of acquisition. Of growing interest is the precise registration of an image to a particular map base, particularly for environmental and land use monitoring; the requirement for accuracy here arises from the potential for using the imagery as evidence in enforcement and other legal actions. For some applications, such as simulated flight animations, ortho-image generation, and slope aspect estimation, the correlation of imagery with digital terrain models is necessary. Registration, uniformly accurate to approximately one pixel, may be required for some projects.

For most image registration and related geometric adjustment issues, the digital image processing community has relied almost exclusively on polynomial warping algorithms. Typically a set of ground control points, or other tie points, are defined to calculate a first or second (occasionally third) order polynomial function to effect the desired transformation. The technique is convenient and computationally efficient. The level of precision noted above is, however, achievable only with narrow field-of-view, nadir pointing sensors. In addition, geographical domains being transformed must be of limited size, and terrain relief must be minimal. If any of these conditions is not met, an alternative geometric correction, or registration, technique must be employed.

The Analytic Sciences Corporation (TASC) has developed several approaches to handle these registration problems. The techniques have been applied to the geometric correction of uncontrolled airborne scanner data, resulting in the generation of "seam-free" mosaic; the correlation of SIR-B synthetic aperture radar imagery with variations in local terrain slope, the development of video flight simulations through mountainous terrain (using SPOT and Landsat Thematic Mapper imagery), and the prototyping of a new ortho-imaging technique.

Some of the methods used are based on the development of sensor models; since sensor position uncertainties are often on the order of hundreds of meters, estimation of the sensing geometry must be accomplished from image (distortion) parameters. Where one cannot model the sensor geometry, a local point-to-point registration

technique may be useful. Based upon an algorithm suggested by Akima, TASC developed this method especially for the terrain correction problem, where geometric distortions are not adequately modeled by readily defined surfaces. Unlike polynomial warping algorithms which provide a best fit in the least squares sense, TASC's Akima warper provides exact co-registration at all user defined tie points.

The purpose of this paper is to examine the limitations and types of error associated with these various methods of geometric correction and registration. The magnitude of the error to be expected in employing these techniques will depend on the accuracy with which ground control points have been mapped, the accuracy with which they can be determined from the reference material, and the accuracy of any digital terrain models being used. All of these error sources will be discussed, and some practical observations regarding these factors will be offered. Included will be issues associated with the extraction of terrain models from stereo imagery. Finally, some guidelines for use in selecting the most appropriate, cost-effective method for specific types of application problems will be presented.

UNE METHODE DE DETECTION DU RESEAU HYDROGRAPHIQUE A PARTIR D'UN MODELE NUMERIQUE DE TERRAIN

A METHOD FOR THE EXTRACTION OF A HYDROGRAPHICAL NETWORK FROM DIGITAL ELEVATION MODEL

Philippe Guillotel *, Réjean Simard **, Béné Goze ** et Christian Roux *

* Groupe Traitement d'Images
Département Mathématiques et Systèmes de communication
Ecole Nationale Supérieure des Télécommunications de Bretagne
B.P. 832, 29285 BREST Cedex, FRANCE

** DIGIM
1100, Boulevard René Lévesque Ouest
Montréal, Québec
CANADA, H3B 4P3

RESUME

Le présent article a pour but de proposer une méthode pour l'extraction du réseau hydrographique à partir d'un modèle numérique de terrain (M.N.T.).

La méthode proposée repose sur un traitement hiérarchique à deux niveaux, effectuant une préclassification des pixels par analyse locale du M.N.T., suivi d'un chaînage des segments du réseau.

Les résultats obtenus, bien qu'incomplets, montrent qu'il semble possible de détecter le réseau hydrographique avec une intervention minimale de l'opérateur, une précision améliorée et un coût plus faible.

SUMMARY

This article proposes a solution for the extraction of a hydrographical network from digital elevation model (DEM) data.

The method (taking into account DEM errors) is based on a two levels hierarchical process, providing a classification by local analysis of DEM, followed by a linking performed on the detected segments.

The results, although not completed, are promising because, with a small operator interference, the network is obtained with a precision which seems to be very accurate and at a reduced cost.

1-INTRODUCTION

Avec l'avènement des techniques de télédétection, et les images du satellite SPOT, il est possible de réaliser semi-automatiquement des cartes topographiques à l'échelle 1:50 000 (Rochon, 1985). Le tracé du réseau hydrographique est une des parties manuelles, et son automatiser devrait engendrer des coûts et délais de production plus faibles. La méthode présentée dans cet article a été développée dans cette perspective, par ailleurs, elle pourrait être aussi avantageusement utilisée par les hydrographes (estimation des réserves en eau, surveillance et contrôle de l'évolution du réseau, irrigation, ...). Ces multiples applications expliquent le fait que la détection automatique du réseau hydrographique ait fait l'objet de plusieurs études. Les approches existantes peuvent se regrouper en deux familles, selon qu'elles s'intéressent:

- à la classification d'éléments linéaires directement dans les images, suivie d'une classification permettant de reconnaître les segments appartenant au réseau.

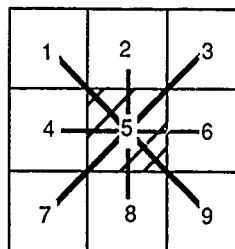
- ou à l'analyse locale du M.N.T., qui repose sur l'étude du voisinage d'un point en vue de sa classification dans le réseau.

Toutes ces méthodes sont efficaces sur un M.N.T. idéal. Il semble pourtant difficile de les appliquer à un M.N.T. pratique, contenant des erreurs.

La méthode présentée ici repose sur un traitement en deux étapes. Dans un premier temps nous introduisons une technique de détection des extrêmes locaux sur un M.N.T. (paragraphe 2), puis un traitement pour le calcul du déversement (paragraphe 3). Dans le paragraphe 4, on présente la méthode de chaînage des segments détectés qui est guidée par un opérateur et tient compte des imperfections du M.N.T. Enfin on évalue les performances de cette méthode en présentant des résultats de traitement sur un M.N.T. obtenu sur un site de Malaisie.

2-CALCUL DES EXTREMA LOCAUX SUR UN M.N.T.

La méthode présentée ici repose sur l'étude du voisinage 3*3 de chaque pixel (voir Figure 1). Le pixel 5 étant le pixel étudié, on calcule les extréma locaux de la façon suivante (Jenson, 1985):



-Figure 1: Liaisons possibles du voisinage 3*3 d'un point-
Possible links in a 3*3 neighborhood

Si les deux extrémités d'une liaison symétrique (1-9; 2-8; 3-7; 4-6) sont d'altitudes supérieures ou égales au point central (point étudié), alors celui-ci sera classé comme minimum local, si, en revanche, les deux extrémités sont d'altitudes inférieures ou égales, alors le point est classé comme maximum local.

Il faut noter que puisque les conditions sont larges, un point peut être classé comme minimum et maximum à la fois. Dans ce cas, on privilégie les minima, c'est à dire que le point sera pris comme minimum, et non comme maximum. Ceci est dû au fait qu'un cours d'eau se présente, dans un M.N.T., sous la forme d'un talweg (profil type "V"), donc comme un minimum local. Ainsi être minimum local est une condition nécessaire pour être cours d'eau, donc dans le doute il vaut mieux opter pour un minimum et ainsi ne pas perdre d'information.

3-CALCUL DU DEVERSEMENT

3.1 Orientation des pentes:

Le but de ce traitement est de calculer l'orientation des pentes suivant 8 directions possibles:

1 : Est; 2 : Nord-Est; 3 : Nord; 4 : Nord-Ouest; 5 : Ouest; 6 : Sud-Ouest; 7 : Sud; 8 : Sud-Est; on rajoute le cas 0 où la pente est nulle.

Pour réaliser ce calcul, on effectue au préalable un filtrage passe-bas du M.N.T., afin de lisser le "relief", et supprimer les erreurs grossières.

Ensuite on localise les points singuliers (points d'altitude inférieure à tous ses voisins, c'est-à-dire les trous), à l'aide de la méthode précédente par exemple.

Enfin, on calcule les valeurs des pentes à l'aide de la méthode des différences finies, ces valeurs permettent finalement de déterminer l'orientation de tous les points du M.N.T. Cette méthode a été développée par Djelali Benmouffok (thèse à paraître, INRS Montréal).

3.2-Calcul du déversement:

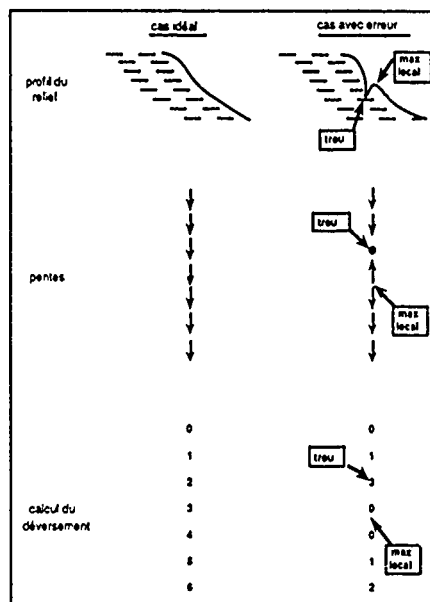
A partir de l'orientation des pentes, on a développé une méthode pour calculer le déversement. Pour se faire, on part du fait qu'une direction de pente, à partir d'un point, nous donne le point dans lequel celui-ci se déverse.

Ainsi, en suivant les pentes, et en incrémentant la valeur de chaque nouveau point dans lequel on se déverse, on peut calculer le déversement du réseau.

Par exemple, un sommet aura une valeur de déversement nulle, alors qu'un ravin (ou un cours d'eau) aura des valeurs de déversement élevées, et ce, d'autant plus qu'on se rapproche de son extrémité.

Malheureusement, la présence des erreurs sur le M.N.T. fausse les résultats. En prenant pour illustration la Figure 2, on s'aperçoit que les valeurs élevées correspondent effectivement au cours d'eau, mais par contre que des valeurs faibles n'impliquent pas nécessairement que ces points n'appartiennent pas au réseau.

Cependant ce calcul permet d'avoir une condition suffisante pour avoir un cours d'eau, à savoir que des valeurs de déversement très grandes impliquent que ces points appartiennent au réseau.



- Figure 2 : mise en évidence des problèmes dus aux erreurs sur le M.N.T. dans les calculs de déversement-
Presentation of some problems due to errors on DEM in the overflow computing

4-APPLICATION A LA DETECTION D'UN RESEAU HYDROGRAPHIQUE A PARTIR D'UN M.N.T.

4.1-Préclassification des pixels:

Dans cette partie nous allons utiliser les diverses informations apportées par les deux calculs précédents, tout en tenant compte du fait que le M.N.T. possède des erreurs.

L'application de ces deux traitements va nous permettre de classer les pixels dans trois classes distinctes:

- la classe 1, contiendra tous les pixels qui appartiennent sûrement au réseau.
- La classe 2, contiendra tous les points qui appartiennent *peut-être* au réseau.
- La classe 0, les points qui n'appartiennent pas au réseau.

Les critères de classement ont été définis comme suit:

- Pour appartenir à la classe 1, un point ne doit pas être un maximum local et doit avoir une valeur de déversement supérieure à un seuil (*SEUIL1*), ou il doit être un minimum local et avoir une valeur de déversement comprise entre le seuil précédent et un autre seuil plus faible (*SEUIL2*).
- Pour appartenir à la classe 2, il doit être minimum local (sans vérifier les conditions précédentes).
- La classe 0 comprendra tous les autres points.

Il reste alors à fixer les seuils; dans un premier temps ceux-ci sont choisis au vu des valeurs de déversement obtenus sur l'image complète. Le *SEUIL1* est choisi élevé afin d'être certains que les points de la classe 1 seront effectivement des points du réseau, le *SEUIL2* quant à lui est pris plus faible pour tenir compte des cas ambigus.

La Figure 3 présente le résultat de classification des pixels dans un fenêtre de 32 par 64 extraite du M.N.T. d'un site de Malaisie.

4.2-Chainage des segments:

Les résultats de l'étape précédente nous donnent des segments du réseau entre lesquels se trouve des points de la classe 2 (appartenant peut-être au réseau). La suite du traitement va donc consister à relier ces segments au travers de ces points "ambigus".

Actuellement ce processus est réalisé avec l'aide d'un opérateur, celui-ci donne l'extrémité d'un segment et la direction de la connexion à effectuer. L'algorithme vérifie si on peut réaliser une connexion. Celle-ci n'est possible que si en parcourant "l'image" dans le sens donné, ou un sens compatible (c'est à dire qui ne soit pas de direction opposée à celle donnée), à partir du point désigné et en suivant des points de la classe 2, on arrive à un autre segment (classe 1).

On impose de plus la contrainte que pendant tout le trajet de connexion, l'altitude soit constante ou décroissante.

Examinons le cas où l'altitude viendrait à augmenter en cours de liaison. On décide alors de corriger l'altitude de ce point (correction limitée à la précision donnée sur le M.N.T. de ± 3 mètres), si cette correction ne suffit pas, la liaison est dite impossible et on passe à un autre segment, sinon on continue le processus de liaison de ce segment.

4.3-Résultats:

Cette méthode a été testée sur un site de Malaisie pour lequel un M.N.T. était disponible. L'image était de taille 128*128 pixels, le paysage de type accidenté, sous couvert végétal de type forêt tropicale, situé dans la région de Nami. L'altitude varie de 0 à 460 m.

Les résultats obtenus (voir Figures 4 et 5) ont donné une reconstruction complète du réseau, avec des cours d'eau non représentés sur la carte, mais conformes à l'allure topographique du site.

En revanche, on pouvait s'attendre à une efficacité réduite pour les faibles altitudes, c'est ce qu'on peut constater sur la partie gauche de l'image (voir Figure 5), pour laquelle le relief est beaucoup plus "doux" (pentes faibles, altitudes faibles). Dans de telles configurations, l'utilisation de l'information radiométrique devrait permettre d'améliorer les résultats.

D'autre part, la précision du tracé semble satisfaisante, mais demande à être évaluée de manière systématique.

5-CONCLUSION

Cette méthode donne de bons résultats sur le site étudié. Elle apporte une première réponse au problème de la détection du réseau hydrographique à partir d'un modèle numérique de terrain pratique (entaché d'erreurs). Si elle fait intervenir un opérateur, sa tâche est simplifiée et réduite. On peut donc s'attendre à une réduction du temps de réalisation des cartes topographiques et une diminution probable des erreurs (notamment des oublis de structures).

Cependant avant de généraliser la portée de ces résultats, une expérimentation systématique reste à faire. Plusieurs améliorations sont envisagées, en particulier l'automatisation complète du processus (y compris éventuellement la détermination des seuils), l'utilisation complémentaire de l'information radiométrique et le développement de système à base de connaissances.

REFERENCES

1. ROCHON, G., "Cartographie topographique à l'aide d'images des satellites SPOT", Association Québécoise de Télédétection, Télédétection et gestion des ressources : l'aspect opérationnel, Vol. V, Bernier, Lessard et Gagnon, pp 367-380, 1985.
2. JENSON, S.K., "Automated derivation of hydrologic basin characteristics from digital elevation model data", Proc. auto-carto 7, digital representations of spatial knowledge, Washington D.C., pp 301-310, March 11-14, 1985.
3. JOHNSTON, E.G. and ROSENFELD, A., "Digital detection of pits, peaks, ridges, and ravines", IEEE transactions on System, Man and Cybernetics, pp 472-480, July 1975.
4. CARROL, R., "Automated gully delineation using digital elevation data", Technical paper, 49th annual meeting ASP. ACSM-ASP convention, Washington D.C., American Society of Photogrammetry, pp 144-151, 13-18 march 1983.
5. RIAZANOFF, S., CERVELLE, B., CHOROWICZ, J., "Ridge and valley line extraction from digital terrain models", Int. J. Remote Sensing, Vol. 9, n°6, pp 1175-1183, 1988.

```

00000000022021222200000022200000000000021120002022111100020000200
00000000002212222000000020200000000002210222002211220002122222220
0000000000212002200000022222000000001100022222122002220222200000
0000000000010002200000222020000002212000022221200000000022000000
000000000002101200002000020000000002000000002210200000000200000000
0000000000021212000220002000000000220000000021200000000000000000
0000000000002110000202200000000000000000000000000000000000000000
22200000000010002222222200000020000000000001220000000020220000
02122000200001021112000022000002202000000200001220000000020220020
00202222200001211200000002000002222000000220022000000000022200220
02000021000021100000000002002221122000000022212000000000021022220
22000002120212000000000002222000112222222111120000000000212022220
000000022121220000000002222000222200022200022000000002120000000
000000000111200000000002222020002200000000220000000021100000000
000000000212000000000002200000002220000000020000022210200000000
00000000001200000000002200000000200000000222000021122000000000
0000000000200000000022222000000000000000022212000211222000000000
00000000222000000002020000000000220022211112000112000000000000
00000000222000000000222000000000020221111222201100120000000002000
2000002221200022000220000000000022222000002200200000000022200
00000222221202222002100000000000002200000000021200000000022000
0202222000122200000120000000000002202200000000212020000000220000
222200000001220000012000000000002220000000000102202200001200000
1220000000212200001020000000000022222200000000002220211112200000
200000000021022220022000222000000222002000000002022111102000000
0000000002120022201220002220000122200200000000200221202212200222
00000002212000221120000212000001220000000000000012000021111222
0000000222000022002000120000001120000000000000011220000222122
00000002220000222022202100000002120000022222222111000000000000
222000002200000002200102000021222222000000000021222200000000
000000000000000000011221220000211200000000020000002122000000000
00000000000000000221221112200021100000000022220022100000000000

```

- Figure 3: résultats de la classification dans une fenêtre de 32 par 64 pixels -
Classification results in a 32 by 64 pixel window

ERROR CORRECTION OF DIGITAL ELEVATION MODELS
PRODUCED BY AUTOMATIC MATCHING OF DIGITAL STEREO IMAGES

J.A. Ostrowski and D.C. He

HORLER INFORMATION INC.
Suite 704, 116 Albert Street
Ottawa, Ontario, Canada K1P 5G3

ABSTRACT

Digital Elevation Models (DEMs) derived automatically from satellite imagery are distorted by three types of errors. Geometric correction errors of between 0.5 and 1 pixel result mainly from the inaccuracies in ground control point location on the map and in the image. Stereo matching errors are slightly above 0.5 pixel for real satellite images. Gross errors (blunders) can exceed several pixels. The new approach to blunder detection and removal, combined with the traditional thresholding and filtering, is capable of removing most of the blunders. To produce blunder-free DEMs, the manual verification and editing stage is still needed.

Keywords: Digital Elevation Model, automatic derivation, errors, blunders

1. INTRODUCTION

Digital Elevation Models (DEMs) contain elevation values of terrain in gridded digital form, and play an important role in map production and many other applications. Extraction of DEMs from digital stereo data by computer correlation (or matching) of image pairs can be now implemented in production systems. This is due to a significant reduction of manual labour required, in comparison to traditional stereoplotting, particularly when processing satellite stereo image pairs.

An automatically produced DEM is distorted by three types of errors. The first type are the geometric correction errors, resulting from inaccuracies of the transformations between the image and ground scene coordinates. The second type of errors are usually small, random errors in determining the exact disparity of the two images, and in interpolating the results to a regular grid. Finally, the third type of errors are blunders, resulting from the matching of the wrong features by the stereo matching

algorithm, or from the failure to detect any matches. In the following, we discuss briefly the magnitude of the three types of errors and possible ways of minimizing their effects in automatic derivation of the DEMs.

2. GEOMETRIC CORRECTION ERRORS

Geometric correction errors originate during the transformation of image to ground scene coordinates. There are three approaches to this task: (i) the transformation function can be derived by comparison of the image and ground scene coordinates of the selected group of Ground Control Points (GCPs); (ii) the orbit and attitude of the satellite can be described by polynomials in time, based on available satellite telemetry parameters and a few GCPs, and then used to determine the corrected position of the pixels; (iii) the satellite orbit and attitude can be described by a physical model, the parameters of which are determined from very few GCPs per strip of scenes.

In the first approach about ten GCPs are needed for one SPOT scene or Landsat TM quarter-scene. The transformation is then described by a low degree (three to five) polynomial in image or ground scene coordinates, and its coefficients determined by least squares fitting. This method, used in most of the commercial image processing systems, has a serious drawback of requiring a relatively large number of GCPs. In well mapped areas it is possible to acquire enough GCPs, although at a cost of long operator time. In poorly mapped areas the cost of GCP acquisition can be very high. Also, a large number of not very accurate GCPs introduces errors in the transformation polynomials. In our tests with an eastern Ontario/western Quebec SPOT PLA scene pair, the use of 32 and 25 GCPs from 1:50,000 maps in the third degree polynomial transformation resulted in residual errors of almost 9 metres in both line and pixel directions. This was only slightly worse than the limit imposed by the SPOT stereoscopic accuracy of about 6 m in X and

Y and 3.5 m in Z, for the maximum stereo effect corresponding to $\pm 27^\circ$ degree viewing angle (Rodriguez et al., 1982).

The second approach is based on expansion of orbital and attitude parameters into low degree polynomials in time. It requires far fewer GCPs to determine the coefficients describing the required transformation, about ten for an area of 500 by 500 km (Guichard, 1985), giving accuracy of 3.5 m for simulated SPOT PLA data. This value did not include GCP location errors.

Similarly the third approach, orbit and attitude physical modelling and integration in time, allows the continuation of the model along a strip of scenes. To determine coefficients of the model, only a few GCPs are needed per strip of ten scenes (Friedmann et al., 1983), yielding an error of about half a pixel (40 m for Landsat MSS data), corrected for GCP error.

The use of GCPs introduces errors due to their uncertain location on the map and in the image. In the context of DEM derivation it poses another difficulty. In off-nadir scenes every point not on the datum surface appears shifted from its map position by a distance (parallax) roughly proportional to the point elevation. Unless all GCPs are exactly at the same elevation, the polynomial or parameter fitting stage includes some of the effects of the parallax. When the general trend of the scene topography matches a low degree polynomial, most of the parallax signal can be removed, and the resulting DEM will be distorted. To avoid such loss of information, the effects of GCP elevation have to be accounted for in the fitting algorithm, or compensated before using GCPs in the fitting stage.

Finally, in applications requiring high absolute accuracy of the determined elevation values, the problem of the reference surface is becoming important (Kaufman and Haja, 1988). The differences between different geoid models (e.g. Clarke 1866 and GRS80) can approach or even exceed total errors of DEMs derived from SPOT images. Unfortunately, a solution to this problem lies outside of the remote sensing community, as it requires the definition of an accurate world reference geoid. On the other hand, due to irregularities of the Earth's shape, some applications may always need corrections for local departures from the reference geoid.

3. ERRORS IN PARALLAX DETERMINATION

There are many methods for automatic stereo images matching and parallax determination. They can be divided into two broad classes: area based matchers and feature matchers. The area matchers seem to perform best in the natural terrain, while the feature detection matchers appear to be well suited to urban and agricultural scenes, with abundant edges.

In laboratory conditions and with known object shape, El-Hakim (1989) was able to achieve location accuracy of 0.05 pixel in X and Y, and depth accuracy of 0.07 pixel. In a similar project, but using synthetic crosses overlain on digital aerial images, Mikhail et al. (1984) achieved the location accuracy of 0.03 - 0.05 pixels in X and Y.

For satellite images the accuracy is at least an order of magnitude poorer. For a simulated pair of pseudo-stereo Landsat images, free of geometric and radiometric distortions, Paine (1986) achieved an accuracy of about 0.4 pixel; additional postprocessing improved the accuracy to about 0.2 pixel, which probably establishes a theoretical limit for satellite images. In tests involving simulated SPOT images, for a realistic case allowing geometric and radiometric differences, Rosenholm (1986) achieved the accuracy of about 0.35 pixel in parallax measurements. The accuracy was two to three times poorer in areas of "bad structure" or with large time difference between images. This is confirmed by results of Cooper et al. (1987): for simulated SPOT data the feature detection algorithm, with postprocessing, gave an error of about 0.7 pixel, while for real TM scenes the error was about two pixels. The errors are usually measured by RMS differences between derived and dense reference DEM; in many cases the accuracy of the reference DEM is comparable to or even lower than the accuracy of the automatically derived DEM. This shows the necessity of evaluation of the automatic DEM derivation systems against accurate, photogrammetric quality dense DEMs.

The errors of interpolation between matched points and a regular grid are most important for feature based matchers, which typically match less than 5% of points. Unfortunately, it is difficult to separate the interpolation errors from other types of errors, although they can be noticed visually (Cooper et al., 1987).

4. GROSS DEM ERRORS

Automatic DEM derivation methods occasionally produce gross errors, or blunders, resulting from matching the wrong features or areas of the two images. The mismatches can result from a variety of conditions, including sensor noise, low scene variance in portions of the images, particularly over water bodies, temporal changes in the scenes imaged at different times, severe relief-induced distortions between the images, and the presence of ambiguities such as occluded areas, clouds, or repetitive patterns. It is very difficult for a matching algorithm to handle all of these situations without error, and therefore means have to be developed to correct gross DEM errors. Most of the emerging production DEM derivation systems operate in an iterative mode with increasing resolution in each step. To prevent the propagation of blunders from coarse to fine

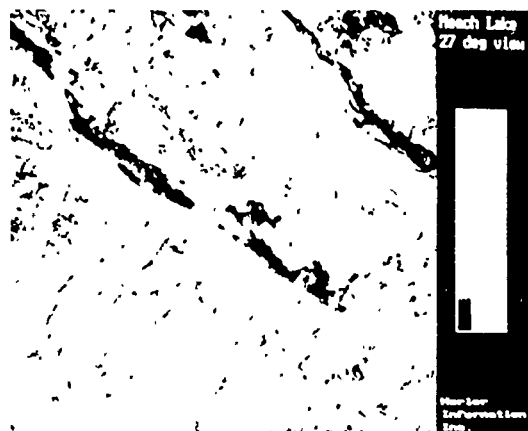


Figure 1a. +27 degree view of a 12.8 by 12.8 km SPOT PLA scene of Meach Lake

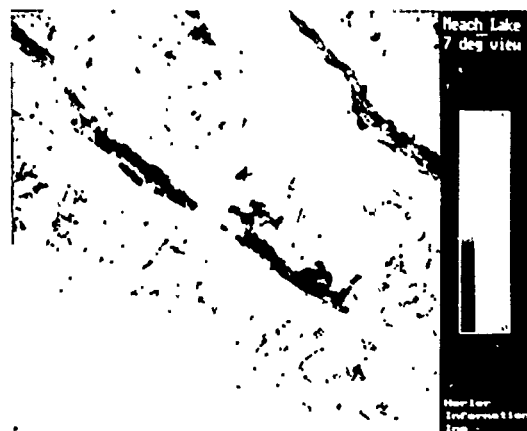


Figure 1b. +7 degree view of the Meach Lake scene

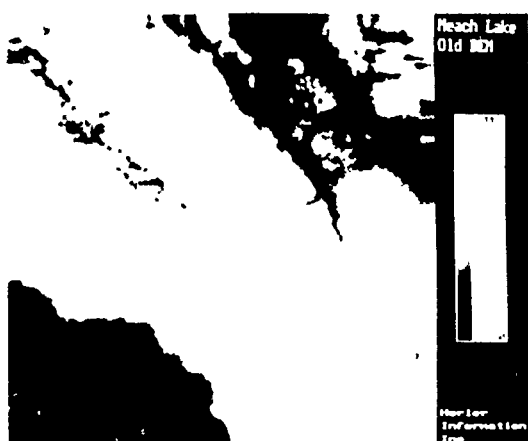


Figure 2. Automatically derived DEM with blunders

resolution, DEM verification and correction has to be embedded into every step of the processing algorithm.

There are many possible approaches to the automatic detection and removal of blunders. They range from a simple thresholding of differences between a point and its surroundings, to complex statistical treatment (Bethel, 1983). All published approaches try to detect blunders by comparing the elevation value at a point with its neighbouring values and deciding whether the determined value is probable. This works in many cases, but due to the great variety of landforms on the Earth's surface there are always real, but improbable features. An algorithm can be fine tuned to accept these features, but this usually means accepting a number of blunders; a restrictive approach reduces the number of blunders, but at a cost of removing some of the real terrain features (e.g. Hannah, 1981). Therefore another approach is needed, detecting blunders by using some physical criteria of their improbability.

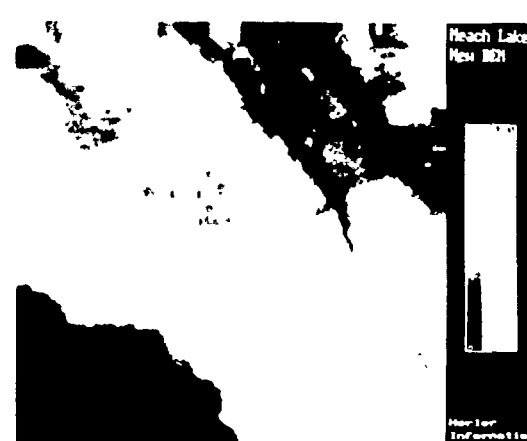


Figure 3. DEM with most of the blunders automatically removed

Our DEM derivation system uses an area based correlator in an iterative scheme, proceeding from the coarsest to the finest resolution. After each iteration the image of determined disparities is screened for its consistency with the viewing geometry under the relaxation of the epipolarity constraint. This detects and removes most of the blunders; the resulting disparity image is further thresholded to remove almost all remaining blunders.

Figure 1 shows two SPOT PLA histogram stretched images of a 1024*1024 scene of Meach Lake (left of centre) just north of Ottawa. The left image is a +27 degree view, showing clouds in both upper corners and some features on the water surfaces. The right image is a +7 degree view. Figure 2 shows a DEM derived in three iterations using only the threshold screening for blunders and heavy filtering. The blunders can be seen as bright and dark spots, concentrated under the clouds and on lake shores. The largest blunder, a depression of 49 meters, occurs in the low, flat area

in the lower left corner of the scene. The RMS difference between our DEM and the elevations of 125 GCPs from the 1:50,000 map of the area was 9.05 m. The inadequacy of using a limited number of GCPs for the evaluation of a deblundering approach is recognized, but we are still in the process of gaining access to a high quality, dense DEM of the test area.

Figure 3 shows a DEM derived using our deblundering algorithm. Most of the blunders are removed, although the largest one still remains, but with error reduced to -28 m. The RMS difference is lowered to 8.83 m, mostly due to the lighter filtering. The smaller filter allowed us to reduce the unprocessed frame surrounding the DEM, which in turn introduced a new blunder on the right side (caused by the river surface patterns seen only in the left image).

5. DISCUSSION

The results of this test, confirmed by similar results for several other scenes, show that our algorithm is effective in removing most of the blunders. The remaining ones survive only because they happen to pass the rejection criteria. This demonstrates that in dealing with real data one has to accept some errors: there are no approaches capable of removing automatically all gross errors without distorting the remaining good results. Therefore, in our opinion, any automatic system has to provide the capability of human verification and intervention; a good system should minimize the number of gross errors, and offer strong support in performing manual verification and correction.

The derivation of a DEM has to be followed by its editing. The obviously required step is lake editing, to assign constant elevation to each lake; this was not applied here to demonstrate the raw blunders. The other possibility would be to mask out lakes and cloudy areas before image matching. Another more difficult step involves river editing, to ensure their flow downstream in the DEM. Also, for some applications, a forest editing step might be needed to correct elevation values for tree height. This would involve classification of multispectral images and acquisition of average tree height data.

To conclude, current systems for automatic derivation of DEMs from satellite imagery are capable of producing models with an absolute accuracy of just below one pixel and relative (internal) consistency better than half a pixel in all three spatial coordinates. It is now possible to remove automatically most of the gross DEM errors, but human verification and correction of the final product is still required. Finally, to produce the highest quality DEMs, techniques for lake, river and forest editing must be implemented in production systems.

REFERENCES

1. Bethel, J.S., 1983. "Surface Approximations and On-line Quality Assessment in Digital Terrain Models", Ph.D. thesis, Purdue Univ., University Microfilms Intl., Ann Arbor, Michigan, 186 p.
2. Cooper, P.R., D.E. Friedmann and S.A. Wood, 1987. "The Automatic Generation of Digital Terrain Models from Satellite Images by Stereo", *Acta Aeronautica*, Vol. 15, No. 3, pp. 171-180.
3. El-Hakim, S.F., 1989. "A Hierarchical Approach to Stereo Vision", *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 4, pp. 443-448.
4. Friedmann, D.E., J.P. Friedel, K.J. Magnussen, R. Kwok and S. Richardson, 1983. "Multiple Scene Precision Rectification of Spaceborne Imagery with Very Few Ground Control Points", *Photogrammetric Engineering & Remote Sensing*, Vol. 49, No. 12, pp. 1657-1667.
5. Guichard, H., 1985. "Etude Mathematique des Possibilités Cartographique de SPOT et Applications Pratiques", in Bernier, Lessard and Gagnon (ed.), *Compte rendu du 5ieme congres de l'A.Q.T.*, Chicoutimi, Quebec, Vol. 5, pp. 355-366.
6. Hannah, M.J., 1981. "Error Detection and Correction in Digital Terrain Models", *Photogrammetric Engineering & Remote Sensing*, Vol. 47, No. 1, pp. 63-69.
7. Kaufman, D.S. and S.R. Haja, 1988. "Extraction of Dense Digital Elevation Models from SPOT Stereo Imagery", *Proc. of IGARSS '88 Symposium*, Edinburgh, Scotland, Vol. 1, pp. 477-478.
8. Mikhail, E.M., M.L. Akey and O.R. Mitchell, 1984. "Detection and Sub-Pixel Location of Photogrammetric Targets in Digital Images", *Photogrammetria*, Vol. 39, No. 3, pp. 63-83.
9. Paine, S.H., 1986. "An Evaluation of Errors in Automated Digital Image Correlation", *Can. J. of Remote Sensing*, Vol. 12, No. 2, pp. 94-102.
10. Rodriguez, V., P. Gigord, A.C. de Gaujac and P. Munier, 1988. "Evaluation of the Stereoscopic Accuracy of the SPOT Satellite", *Photogrammetric Engineering & Remote Sensing*, Vol. 54, No. 2, pp. 217-221.
11. Rosenholm, D., 1986. "Numerical Accuracy of Automatic Parallax Measurements of Simulated SPOT Images", *Can. J. of Remote Sensing*, Vol. 12, No. 2, pp. 103-113.

Least Squares Prediction Using On-Board Data In Bundle Adjustment for SPOT Imagery

Liang-Chien Chen and Liang-Hwei Lee

Center for Space and Remote Sensing Research
National Central University
Chung-Li, Taiwan, R.O.C

ABSTRACT

Due to its high geometric precision and off-nadir looking characteristics, SPOT satellite imagery has a very high potential in three dimensional mapping. The first step in stereo mapping is to determine the exterior orientation parameters for sensors. For SPOT imagery, polynomials are often used to describe the relationship between sampling time and orientation parameters. While the valuable on-board orbital and attitude data are seldom used. This paper is to use on-board data and least squares prediction technique to make the orientation determination in bundle adjustment more reasonable. The main purpose of this study is to minimize the influence of the number of ground control points to positioning accuracy. Accuracy analysis is included and the relationship between accuracy and the number of ground control points is also investigated.

1. INTRODUCTION

Since the launch of SPOT 1 in Feb. 1986, hundreds of thousands images have been collected for various applications. World-wide research efforts have been concentrated on extracting 3-D topographic information, DTM for instance, from SPOT data due to its high resolution and off-nadir looking characteristics.

There are two approaches to use SPOT data in stereo mapping. The first is to use digital data to determine the orientation parameters for sensors, to generate DTMs and ortho images, and to extract the ground features. The second approach is to use analytical plotters and analogue films. The application softwares currently in use on analytical plotters need to be modified to fit the SPOT satellite sampling geometry. After the modification, the operating procedures can follow the way as for central perspective imaging data. The major difference between digital approach and using analytical plotters is that the

different data type being used. The process is thus different. The first step in stereo mapping using SPOT data is to determine the exterior orientation parameters for sensors at different time. In our previous study [1,2], we use bundle adjustment in the orientation modelling, in which third order polynomials are employed to characterize the orbital data and attitude as function of time. The RMS errors for check points were 7.4m, 5.0m, and 6.9m for X, Y, Z coordinates respectively when 28 GCPs were used. In order to minimize the influence of the number of control points to positioning accuracy, we are now introducing the on-board attitude data in the solution. Position-velocity and attitude rate are included in CCT for a SPOT image. Where nine sets of orbital data are available with one minute interval. The scene center sampling time is between set 4 and 5. Seventy two angular velocity sets for roll, pitch and yaw are recorded at 82 scanned line interval for panchromatic band. If those on-board data can be used properly, it might improve the accuracy in determining orientation parameters. That would reduce the parameters in the functional model and the fewer number of GCPs can be expected to yield a higher accuracy. [3]

This paper is to use lower order polynomials, first and second order, incorporating on-board data in the CCT and linear least squares prediction technique to predict the attitude variation. Then introduce those predicted values in the bundle adjustment as constraints.

2. MATHEMATICAL MODEL

2.1 Modified Collinearity Equations

In a SPOT panchromatic scene, we have 6000 lines. Due to its pushbroom scanning characteristics, image projection in sample direction is perspective and is approximately parallel in flight direction. Therefore the modified collinearity equations are

$$x = 0 = -f \frac{m_{11}(X-X_1) + m_{12}(Y-Y_1) + m_{13}(Z-Z_1)}{m_{31}(X-X_1) + m_{32}(Y-Y_1) + m_{33}(Z-Z_1)}$$

$$y \cdot S_y = -f \frac{m_{21}(X-X_1) + m_{22}(Y-Y_1) + m_{23}(Z-Z_1)}{m_{31}(X-X_1) + m_{32}(Y-Y_1) + m_{33}(Z-Z_1)}$$

$$X_1 = X_0 + X_1 \cdot t + \dots$$

$$Y_1 = Y_0 + Y_1 \cdot t + \dots$$

$$Z_1 = Z_0 + Z_1 \cdot t + \dots$$

$$\omega = \omega_0 + \omega_1 \cdot t + \dots$$

$$\varphi = \varphi_0 + \varphi_1 \cdot t + \dots$$

$$\kappa = \kappa_0 + \kappa_1 \cdot t + \dots$$

m_{11} - m_{33} are elements of rotation mat
 S_y : scaling factor

2.2 Observation Equations for On-Board Data

(1) Position-Velocity Equations

We have 9 consecutive orbital data at 1 minute interval, therefore the sampling area is located between no.4 and 5.

The position-velocity data can not strongly constrain a single scene because the sampling time for a scene is about 9 seconds while the recording interval is one minute. However, it still can be used as initial values and as weighted observations in the adjustment.

The observation equations are :

$$\bar{X}_i(t) = X_0 + X_1 \cdot t_i$$

$$\bar{Y}_i(t) = Y_0 + Y_1 \cdot t_i$$

$$\bar{Z}_i(t) = Z_0 + Z_1 \cdot t_i$$

(2) Attitude Rate Equations

The data describe the attitude change for a scene at 82 lines interval as show in Fig.1. The data are more valuable than position-velocity data in positioning, because it covers the whole scene and affluent data, 72 sets, are included and the most important is that the variation is pretty systematic rather than random.

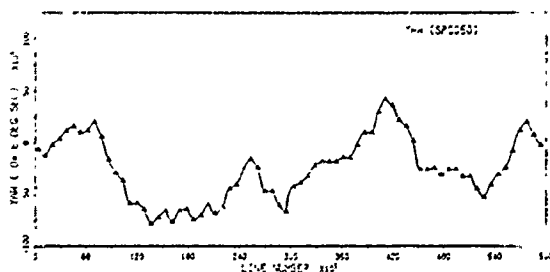


Figure 1 The Sample of Attitude Rate Data.

The observation equations are :

$$\omega'(t) = d\omega/dt = \omega'$$

$$\varphi'(t) = d\varphi/dt = \varphi'$$

$$\kappa'(t) = d\kappa/dt = \kappa'$$

2.3 Linear Least Squares Prediction [4,5]

$$H_{int} = c C^{-1} H$$

$$= [c(PF_1) \dots c(PF_n)] \begin{bmatrix} V & c(P_1 P_2) & \dots & c(P_1 P_n) \\ & V & & c(P_2 P_n) \\ & & \ddots & \vdots \\ & & & V \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$

where

H_{int} : interpolated value.
 c : Covariance matrix between reference pts. and interpolated pts.
 C : Covariance matrix between reference points.
 H : reference value (the difference of recording value and trend)

The covariance function can assume to be Gaussian :

$$W = C_1 \cdot \exp(-C_2 \cdot d^2)$$

where

W : Weight.
 c_1 : filtering parameter.
 c_2 : curve parameter.
 d : distance between points.
 When $c_1=1$, filtering for reference value is performed.

3. CASE STUDY

The stereopair used in the study is located in central Taiwan with base-height ratio 0.57. The other related descriptions are shown in table 1.

| Scene ID | Sampling Rate | Sensor | Incidence | Scene Center |
|----------|---------------|--------|-----------|----------------------|
| SPC049 | 1987, 01, 15 | HRV2 | L 10.4 | N0242958 E1204303 |
| SPC050 | 1987, 01, 16 | HRV1 | R 24.1 | N0242958 E1205840 |

Table 1. General Descriptions for SPC049 and SPC050.

Fifty uniformly distributed and easily recognized points in the stereomodel were selected as control points or check points. The configuration for 28 control points and 22 check points is shown in Fig.2. The ground coordinates for control and check points were digitized from 1/5,000 scale photo base map sheets with a Calcomp 6000 digitizer.

For image points, we directly read the line/sample number of the left image on a image processing system, IDIMS, as image coordinates then used least squares match-

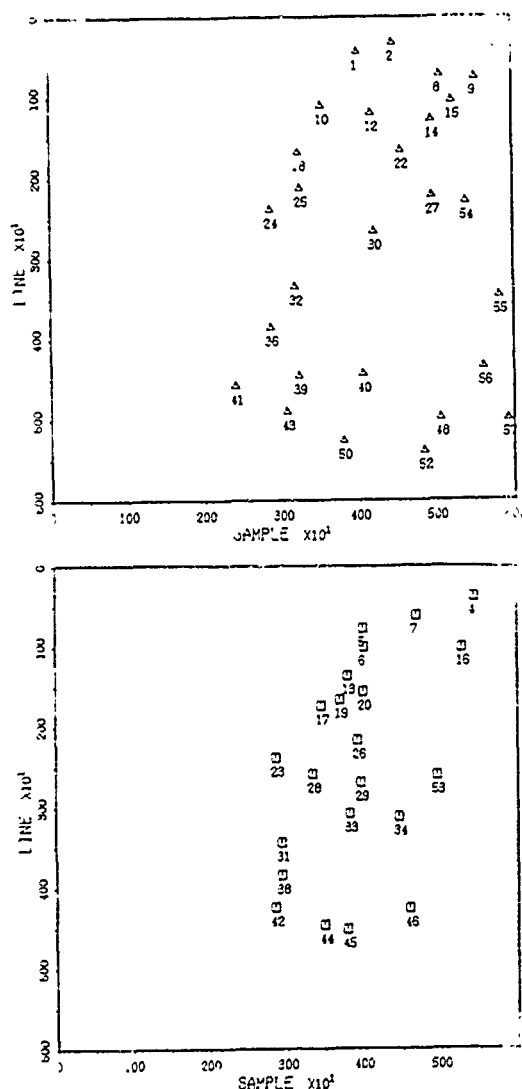


Figure 2 Configuration for 28 Control Points and 28 Check Points.

ing technique to determine the position for the conjugate points in the right image to sub-pixel level.

Before entering the bundle adjustment, we used least squares prediction according to the on-board data to interpolate the attitude data for points of interest. Then a bundle adjustment with additional parameters for SPOT data was developed and executed. We examine the difference between the difference using on-board data or no. We also compare the performances for the first order and the second order polynomials in the functional model for attitude. Fig.3 shows the RMSE of check points for first order attitude model. Fig. 4 shows the RMSE of check points for second order model. The performances of using on-board or no are represented res-

pectively by (a) and (b) in both figures.

4. CONCLUSION

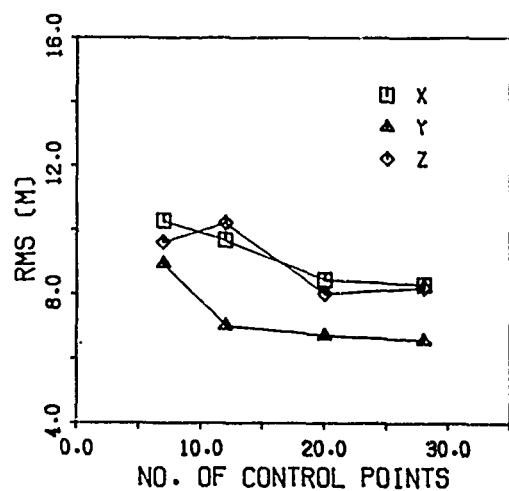
(1). This study is aimed to reduce the number of GCPs and the results indicate that auxiliary use of on-board data are indeed helpful. For lower order attitude functional models and small number of GCPs, least squares interpolation for on-board data can improve the positioning accuracy. It is very significant in the first order case. For the second order functional model, the accuracy improvement is obvious but less than that of the first order. This is because that even without using on-board data, more parameters are introduced in the second order attitude function to compensate the systematic errors to yield better accuracy.

(2). Accuracy improvement by use of satellite data tends to increase when smaller number of GCPs are used. The optimal number of GCPs still needs further investigations to determine, because the number of control points should be considered with its distribution.

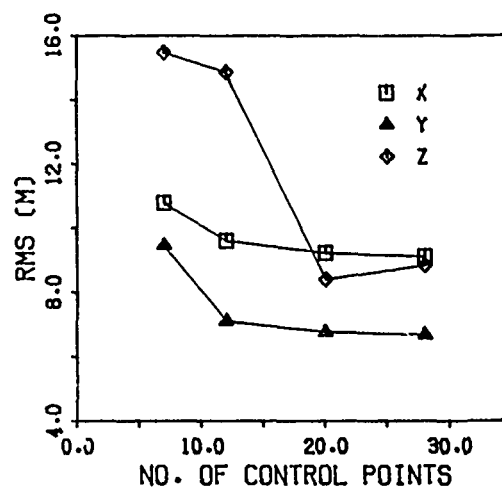
(3). The registration errors between photo base map and SPOT stereomodel are introduced for control and check points in this study. This half - pixel expected error could cause the accuracy potential underestimated. We have an ongoing project to investigate this.

REFERENCE

1. Chen, L. C., Lee, L. H. & Lee, S. C., 1988, DTM Generation Using SPOT Digital Data, XVI ISPRS Congress, Commission III, Kyoto, pp 100-110.
2. Lee, L. H. & Chen, L. C., 1988, Bundle Adjustment with Additional parameters for SPOT Stereopairs, XVI ISPRS Congress, Commission III, Kyoto, pp.III-1-10.
3. Shibasaki, R., Murai, S. & Okuda, T., SPOT Imagery Orientation with Auxiliary Satellite Position and Attitude Data, XVI ISPRS Congress, Commission III, Kyoto, pp. III-125-132.
4. Kraus, K. & Mihail, E. M., 1972, Linear Least Square Interpolation, Photogrammetric Engineering, Vol.38, No.5, pp. 1016-1029.
5. Wu, J., 1988, A Sound Approach for Stereo Three Scanner's Images, XVI ISPRS Congress, Commission III, Kyoto, pp. III-45-51.

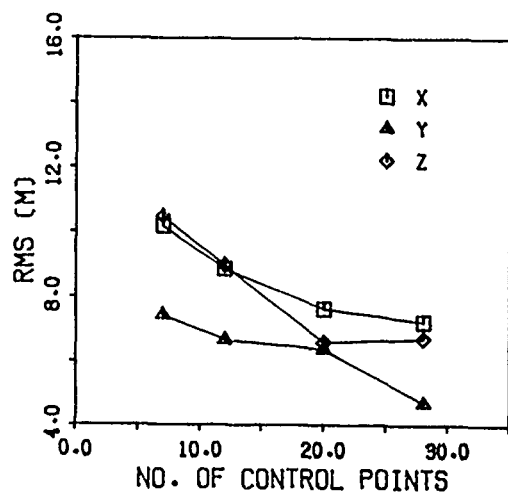


(a) With On-Board Data

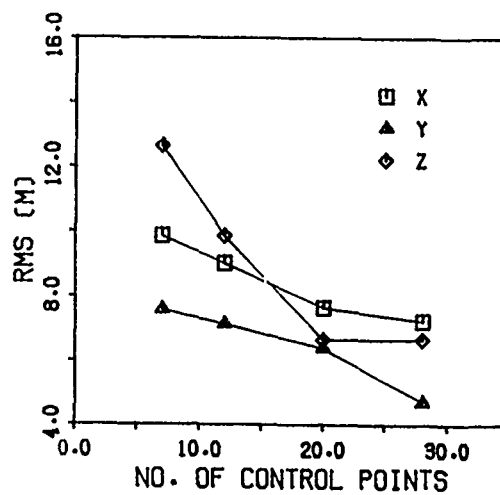


(b) Without On-Board Data

Figure 3. The RMSE of Check Points for First Order Model.



(a) With On-Board Data



(b) Without On-Board Data

Figure 4. The RMSE of Check Points for Second Order Model.

Investigation of slope and aspect effects on image grey values in geocoded SAR images

T. Bayer, R. Winter, G. Schreier

DLR - German Aerospace Research Establishment
Oberpfaffenhofen
Germany

ABSTRACT

The state of a study investigating slope and aspect effects on the backscatter values in geocoded SAR images is given.

In order to achieve detailed results a precise reconstruction of the imaging geometry is required. This requirement is met by the application of image specific Doppler shift functions and orbit polynomials for each pixel of the digital elevation model. The geometric model for this reconstruction of the imaging geometry is given.

Detailed incidence angle calculations are possible and the results are shown by presenting two incidence angle maps. These maps represent two components of the incidence angles: (1) the total incidence angles lying on solid angles, (2) the analysis of the differentiation of the incidence angle in look direction. Some early results of the investigation of incidence angle effects on various landuse classes are used to demonstrate the problems and possibilities of finding empirical backscattering functions depending on incidence angles.

Keywords: SAR Imaging Geometry, Incidence Angle Calculations,
Slope-/Aspect Variations

1. INTRODUCTION

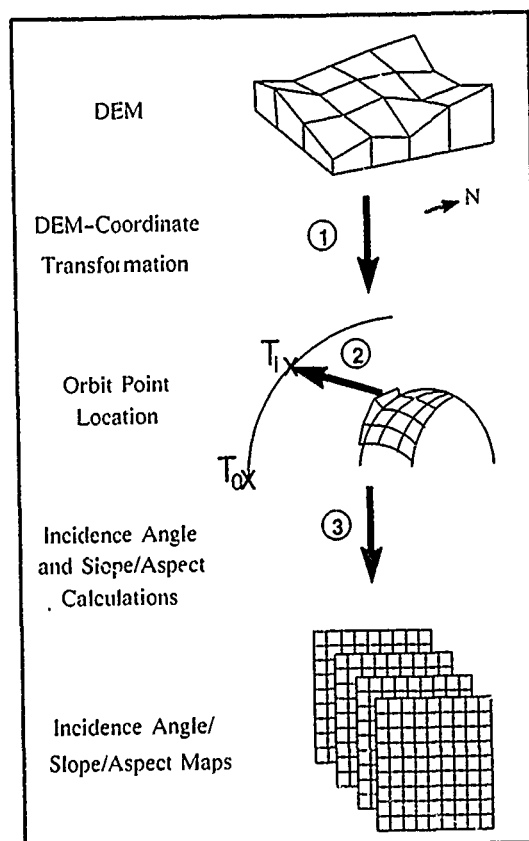
The investigation of relief effects on backscatter values in radar images is a major subject of research recently being conducted in the analysis of spaceborne SAR imagery. Relief effects like slope and aspect variations together with the radar look angle (off-nadir) cause variations in backscatter values for different landuse categories by influencing the local incidence angles of the radar beams. Various approaches have been attempted to describe the relationships between these main parameters of this problem (see references).

In this paper the state of a study is presented that is aimed to investigate slope and aspect effects on spaceborne SAR imagery using as an example a Seasat image of the area of Bonn, FRG. As a prestep to a detailed analysis of incidence angle and aspect variations on backscatter values a precise reconstruction of the imaging geometry is attempted. This reconstruction leads to the computation of two incidence angle maps along with several additional geometrical informations (i.e. slope and aspect parameters). The emphasis of this paper lies on the description of the pixel by pixel algorithms used for these computations.

2. GEOMETRIC MODEL

The use of a digital elevation model (DEM) with the same grid size as the geocoded image is of basic importance for the precise reconstruction of the imaging geometry.

The geometric model for the computation of local incidence angles can be split into the following three steps (see Fig. 1):



- figure 1 -

- 1) Transformation of the coordinates of the DEM into earth centered cartesian coordinates in regard to the surface height (DEM-values);
- 2) Location of the corresponding sensor points in orbit for each earth surface point (pixel in the DEM) using orbit polynomials and an iterative method;
- 3) Calculation of two local incidence angle maps and geometrical parameters via vector geometry regarding surface normal and the azimuth of the look direction.

Within this chapter steps 2) and 3) will be discussed in more detail. The transformation of geodetic or geographic coordinates of the DEM into cartesian coordinates is done using geodetic transformation formulas (Großmann, 1976)

2.1. Earth to Orbit Point Location

For each earth surface point represented by a 3-dimensional vector in the earth centered cartesian coordinate system the corresponding sensor point in orbit is calculated. Inputs into the iterative calculation of the orbit points are:

1) the vector of the earth surface point \vec{p} given as a function of the reference ellipsoid parameters (N,e), latitude (b), longitude (l) and height above the ellipsoid (h):

$$\vec{p} = \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \begin{bmatrix} (N+h) \cdot \cos b \cdot \cos l \\ (N+h) \cdot \cos b \cdot \sin l \\ ((1-e^2) \cdot N+h) \cdot \sin b \end{bmatrix} \quad (\text{Torge, 1975});$$

2) the velocity vector of the earth surface point \vec{p} given as a function of the angular speed (ω_e):

$$\vec{p} = \begin{bmatrix} x_{p^0} \\ y_{p^0} \\ z_{p^0} \end{bmatrix} = \begin{bmatrix} -\omega_e \cdot y_p \\ \omega_e \cdot x_p \\ 0 \end{bmatrix} \quad (\text{Reck, 1988});$$

3) the vector of the sensor point in orbit \vec{s} given as a function of time (i.e. azimuth, i), along with a starting value i_0 :

$$\vec{s} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} a_0 + a_1 \cdot i + a_2 \cdot i^2 \\ b_0 + b_1 \cdot i + b_2 \cdot i^2 \\ c_0 + c_1 \cdot i + c_2 \cdot i^2 \end{bmatrix}$$

Second degree polynomials calculated out of the range line ancillary records of the slant range data proved to be of reasonable accuracy (Crauber et al., 1988).

The basic assumption of the iteration is that for a given ground point the correct orbit point satisfies the SAR Doppler equation. This means the root of a real function must be found. As the derivative of the SAR Doppler equation is known analytically, the Newton iteration algorithm can be used for the iteration

$$T_{n+1} = T_n + DT \\ DT = -F(T) / (dF(T) / dT)$$

where F is the Doppler equation given by

$$\lambda \cdot FD \cdot R/2 + (\vec{p} - \vec{s}) \cdot \frac{\vec{s}}{|\vec{s}|} \cdot \frac{\vec{p}}{|\vec{p}|} = 0$$

as a function of the wavelength (λ), the Doppler reference function (FD), and the range equation (R); DT is the iterated time-step and T_i are the time-dependent points in orbit for specific ground points (Reck, 1988).

During the iteration the following values have to be calculated.

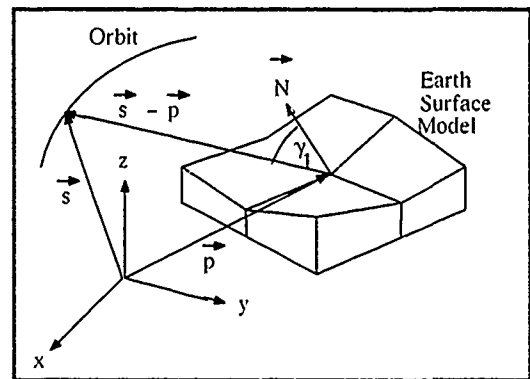
- sensor point vector in orbit (\vec{s}),
- velocity vector of the sensor in orbit ($\frac{d\vec{s}}{dt}$),
- the range equation (R),
- the Doppler shift distribution (FD),
- the Doppler frequency shift (F),
- the derivative of the Doppler function (dF),
- the azimuth (resp. time) step (DT)

Output of the iteration are the azimuth (resp. time) values of the orbit reference polynomials for each earth surface point. These values are calculated in a first major step and are added as an additional layer to the height values of the digital elevation model.

2.2. Incidence Angle Calculation

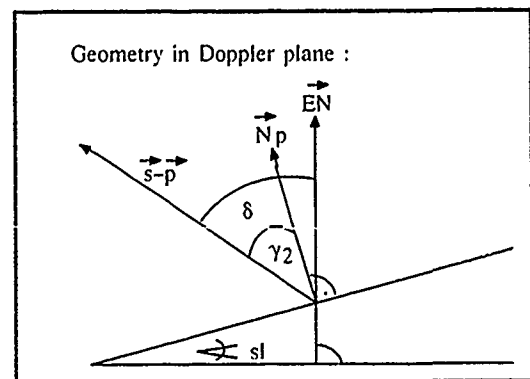
Using these "enlarged" DEM-data along with the coefficients of the orbit reference polynomials as input, two local incidence angles plus slope/aspect values can be calculated for each earth surface point (i.e. for each DEM-pixel).

The incidence angle γ , is the angle between the surface normal (\vec{N}) and the vector in look direction ($\vec{s} - \vec{p}$) (see Fig. 2). A specific value in the resulting incidence angle map therefore represents all incidence angles lying on a solid angle with that same size without regard of the aspect of the corresponding earth surface section.



- figure 2 -

The gradients for each DEM pixel used for the representation of the surface normal $\vec{N} = (-p, -q, 1)$ are calculated by formulas given by Horn (1981). Normally an environment of 8 z-values around the pixel in question is used, at the borders of the DEM more simple approaches are applied.



- figure 3 -

The incidence angle γ_2 is the derivative in look direction of the incidence angle γ_1 , i.e. the angle between the projection of the surface normal into the Doppler plane (\vec{N}_p) and the vector in look direction ($\vec{s} - \vec{p}$). γ_2 can be viewed as the component of γ_1 in the Doppler plane (Rauste, 1988). The geometry is given in Figure 3 where the paper plane is regarded as the Doppler plane.

Instead of computing the projection \vec{N}_p of the surface normal into the Doppler plane, γ_2 is calculated indirectly through the difference of the slope angle in look direction ($\angle sl$) and the angle δ between the vector in look direction ($\vec{s} - \vec{p}$) and the ellipsoid normal (\vec{EN}) at the given ground point. The slope in look direction is given by Horn (1981) with

$$\angle sl = p \cdot \cos \alpha + q \cdot \sin \alpha$$

where α is the azimuth of the look direction.

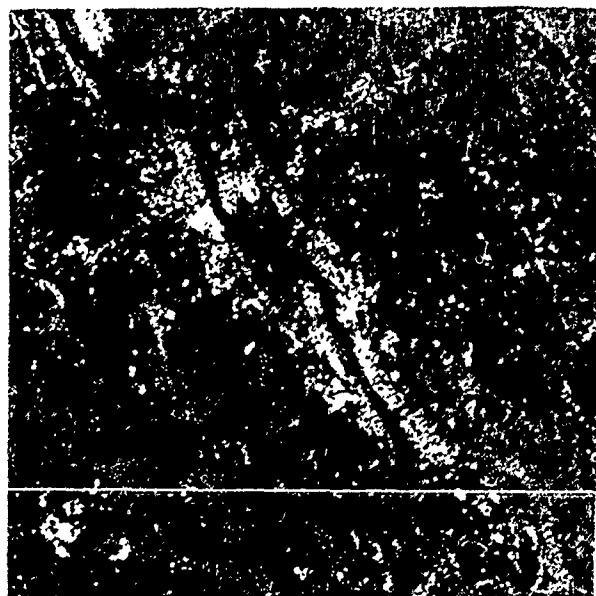
The ellipsoid normal \vec{EN} is given as a function of latitude (b) and longitude (l) by the equation

$$\vec{EN} = \begin{bmatrix} x_{EN} \\ y_{EN} \\ z_{EN} \end{bmatrix} = \begin{bmatrix} \cos(b) \cdot \cos(l) \\ \cos(b) \cdot \sin(l) \\ \sin(b) \end{bmatrix} \quad (\text{Torge, 1975})$$

If the difference $\delta - \angle sl$ is zero or negative the corresponding pixel in the resulting incidence angle map is masked to indicate an area of direct layover.

Additionally the maximum slope and its exposition to the radar beam are calculated.

The digital elevation model, the incidence angle values and the slope and aspect values together with the image grey values of the geocoded SAR-image containing test-site specifications form a multilayer data base as input for further statistical analysis.



- figure 4 -

3. RESULTS AND APPLICATIONS

The procedure introduced above has been applied to the data of a Seasat subimage of the area of Bonn, FRG (see Fig. 4). The subimage shows the Rhine valley at Bonn, to the right of the river there is the mountainous region of the Siebengebirge, to the left the plateau of the Ville.

The image was geocoded at the University of Zurich, Department of Geography, using a digital elevation model with a 25m grid (Meier et al. 1986). The same DEM was used to compute the incidence angle, slope and aspect maps. Greyvalue-coded representations of the incidence angles γ_1 and γ_2 are presented in the Figures 5 and 6.

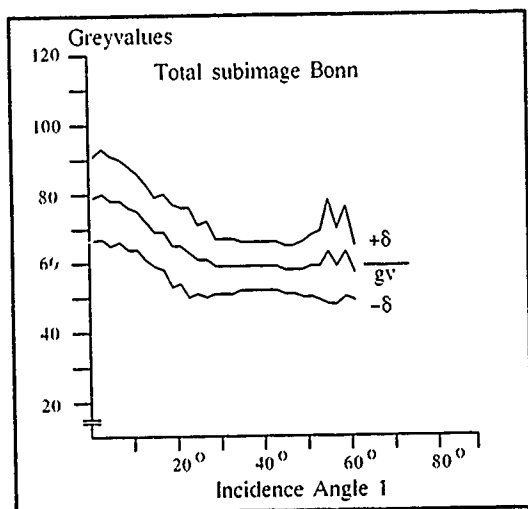


- figure 5 -

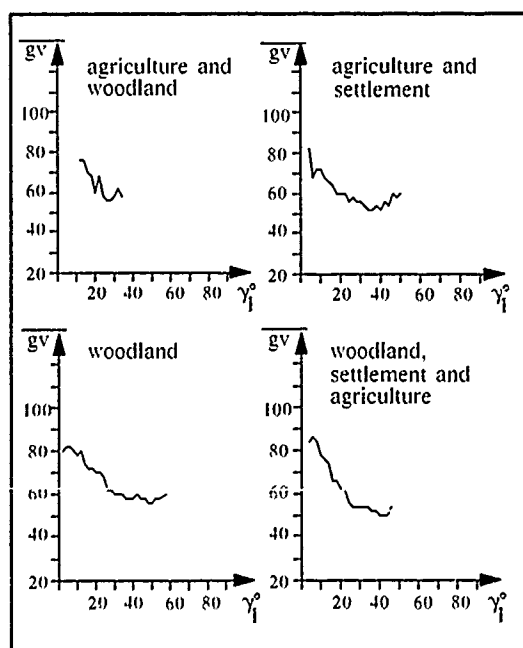


- figure 6 -

The presentations given here are a reproduction of color-coded photographic products of the incidence angle maps. High angle values (up to about 80°) are presented in black and dark grey, incidence angles of around 23° (i.e. flat areas) in grey and smaller values in light grey tones. The areas of direct layover in incidence angle map 2 (see Fig. 6) are marked by black values. These areas appear on many south-west exposed slopes due to the direction of the incident Seasat radar beam. At first sight only little differences are visible between the two incidence angle maps. These differences are mainly situated within the mountainous regions of the Siebengebirge. The use of the incidence angle maps lies in further statistical analysis of the behaviour of image grey values of different test sites for different incidence angles and aspects to the radar beam. By the application of improved algorithms on the values representing incidence angles γ_i the possibility to compute indirect layover areas as well as direct and indirect shadow areas is given. Such algorithms were introduced by Kropatsch et al. (1988) for example.



- figure 7 -



- figure 8 -

In an earlier state of this work incidence angles γ_i were plotted against grey values of the geocoded subimage. Figure 7 shows the results giving mean grey values and the standard deviations for the whole subimage, thus giving only the general behaviour.

For different test sites the results of the plots show special behaviour for different landuse categories (see Fig. 8).

To obtain more detailed results on the behavior of backscatter values with the aim of describing empirical relations for special landuse categories the further steps will be a refined definition of test areas in regard to different landuse categories and a more detailed statistical analysis of the relationship between image grey values and incidence angles and image grey values and aspect to the radar beam. Results of these studies will be used to model specific empirical backscatter models in order to correct geocoded SAR images for relief influences.

4. REFERENCES

1. CIMINO, J.B., ET AL., *Multiple Incidence Angle SIR-B Experiment Over Argentina: Mapping Of Forest Units*, IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-24, No. 4, 1986.
2. CRAUBNER, A., ET AL., *Calculation Of A Reference Function For Orbit Position From Orbit Ancillary Data*, Internal Investigation Notice, DLR, WT-DA-FE, 1988.
3. DOMIK, G., ET AL., *Dependence Of Image Grey Values On Topography In SIR-B Images*, Internat. Journal of Remote Sensing, Vol. 9, No. 5, 1988.
4. GROBMANN, W., *Geodätische Rechnungen und Abbildungen in der Landesvermessung*, Konrad Wittwer, 1976.
5. GUINDON, B., ET AL., *Relief Effects And The Use Of Terrain Models In SAR Image Processing*, Proc. of the 3rd Seasat-SAR Workshop at Frascati, Italy, 1980.
6. HINSE, M., ET AL., *Radiometric Correction Of C-Band Imagery For Topographic Effects In Regions Of Moderate Relief*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 26, No. 2, 1988.
7. HORN, B., *Hill Shading And The Reflectance Map*, Proc. of IEEE, Vol. 69, No. 1, 1981.
8. KROPATSCH, W., ET AL., *The Generation Of SAR Layover And Shadow Maps From Digital Elevation Models*, IEEE Trans. on Geoscience and Remote Sensing, accepted 1988.
9. MEIER, E., ET AL., *Investigation In The Field Of Spaceborne SAR Image Rectification*, Technical Note To German ERS-1 Offline Ground Segment, 1986.
10. RAUSTE, Y., *DEM-Based Image Processing Methods For SAR Images*, Proc. of ISPRS, Kyoto, 1988.
11. RECK, M., *Location Of Image Points Using Doppler And Orbit Reference Function*, Internal Investigation Notice, DLR, WT-DA-FE, 1988.
12. TEILLET, P.M., *Slope-Aspect Effects In SAR Imagery*, Canadian Journal of Remote Sensing, Vol. 11, No. 1, 1985.
13. TORGE, W., 1975, *Geodäsie*, Walter de Gruyter.
14. ULABY, F.T., ET AL., *Microwave Remote Sensing - Active And Passive*, Vol II, Addison-Wesley, 1982.

Quantitative Analysis and Classification of SeaMARC II 12 kHz Side-scan Sonar Data.

THOMAS BECKETT REED IV

Hawaii Institute of Geophysics, Honolulu, Hawaii

Digital image processing techniques are applied to side-scan sonar images collected with the 12 kHz SeaMARC II system to effect total radiometric and geometric rectification. Subsequent bottom-up region growing, followed by texture and fractal dimension analysis yields feature vectors on the basis of which regions of limited image heterogeneity can be classified. The region growing technique allows second and higher order statistics to be evaluated over features with irregular boundaries, without incorporating the statistics of the boundary or the neighbors, with a region resolution on the order of twice the pixel length (80m). The use of second order statistics makes the techniques insensitive to absolute intensity level, and yields relative measures of surface roughness.

1. INTRODUCTION

Side scan sonars have disclosed the location and, to some extent, the morphology of seafloor features but have provided little of the type of quantitative information necessary for lithologic identification. This quantitative imprecision, in contrast to the successes presently being enjoyed by the terrestrial remote sensing groups, stemmed chiefly from technical hardware problems resulting in a lack of seafloor image fidelity. Now, however, the application of objective digital image processing techniques to images of superior quality and uniformity makes possible a more quantitative approach in the analysis and interpretation of synoptic seafloor imagery.

1.1 The SeaMARC II System

SeaMARC II combines a conventional side scan sonar with a bathymetric mapping system in a single unit towed at depths of 100 m or less at speeds up to 10 knots (5 m/s). In water depths greater than 1 km, the system produces 10 km wide data swaths, permitting 100% coverage of over 3000 km of seafloor per day. A complete system description is given by Blackinton *et al.*, [1983].

Two parallel, inclined arrays are mounted on each side of the SeaMARC II towfish. The port arrays operate at 11 kHz and the starboard at 12 kHz to minimize cross-talk. By assuming a nominally flat bottom, and calculating the rate at which the outgoing pulse will have swept the bottom for cross-track distances from nadir to 5 km athwartships, the returning signal is divided into 1024 unequal intervals of time, each representing a 5-m-wide swath of the seafloor. Sampling rates are high for the near-nadir pixels and decrease in proportion to the cosine of the grazing angle. If the seafloor is in fact flat across-track, then this process will produce an image that is geometrically correct in horizontal range, but for ray bending effects. Cross-track topographic variations will, however, result in image distortion. Specifics of this type of distortion, and remedies for it, are presented in section 2.

The SeaMARC II bathymetric information is acquired through the same transducers as the side scan data but is processed with different hardware and software. On transmission the transducer pairs on each side of the towfish are driven in parallel. On reception, each row in the pair is sampled independently. Therefore any signal incident upon the transducer at any angle off normal to the transducer face is detected at the two rows with a different phase lag, from which the depression angle θ of the reflector is calculated. By measuring the round trip travel time and assuming a sound speed of 1500 m/s, the slant range R to the reflector can be calculated. These values for R and θ are converted to across-track distance and depth for each reflector and contoured to produce a bathymetric map. Absolute accuracy of 2-3% of the water depth is nominal; relative accuracy is significantly better.

2 PREPROCESSING CORRECTIONS APPLIED TO SEA MARC II DATA

A side scan sonar image is a two-dimensional display of pixels, each with an associated intensity, which attempts to model a physical realiza-

tion of a four-dimensional process, namely, the interaction of sound with the seafloor. The dimensions of the process are the three Cartesian coordinates of space, which give the position and orientation of any reflector with respect to the sonar, and the fourth dimension of acoustic character (acoustic impedance and roughness). The image is an approximation of the bottom with various sources of errors. Preprocessing corrections should produce an image that is geometrically and radiometrically correct. Being geometrically correct implies that features are in their correct locations and represented by the same spatial distribution of picture elements irrespective of the slant range at which they are imaged. Being radiometrically correct implies that all contributions not indicative of actual changes in bottom acoustic character have been removed. Only when these two criteria are met can meaningful subjective or quantitative interpretations be conducted.

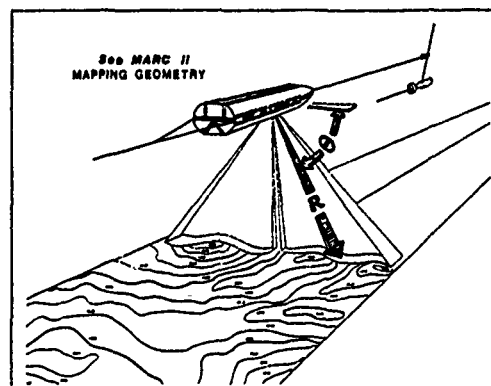


Fig 1 SeaMARC II deployment configuration. Towfish depth ca. 100 m

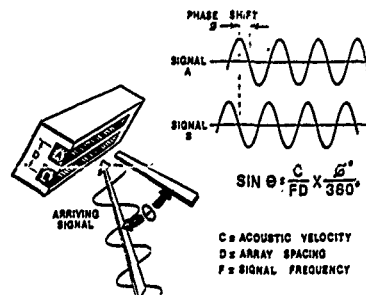


Fig 2 SeaMARC II phase interferometric bathymetry system.

2.1 Radiance Transformations: Background Subtraction

Three radiometric artifacts are common to many side scan images (Figure 3). The first is the irregular band of high intensity pixels nearest to the ship's track (SeaMARC II plots strong reflectors as dark and shadows as light, so the images appears as negatives). The high intensity of these pixels is due to the contribution from near normal incidence specular and subbottom returns, which will only occur within the first few degrees from nadir. The second artifact consists of lines of high-intensity pixels parallel to the ship's track which represent surface reflections of the first bottom echo. The lines nearest the ship's track are peg-leg (down to the bottom, up to the surface, and back down to the towfish again) multiples. Similar dark linear features roughly parallel to the peg-legs but at the outside edges of the image are the first bottom multiples.

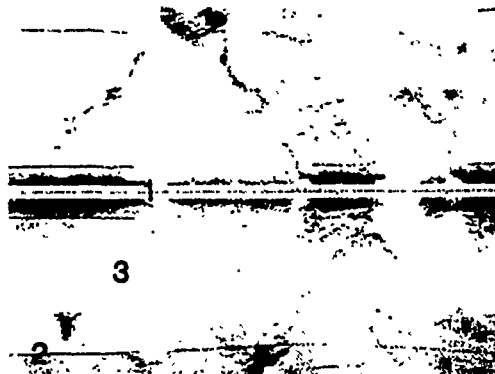


Fig. 3. SeaMARC II side scan image, containing three types of errors: (1) high amplitude near nadir specular returns; (2) peg-leg and first bottom multiples; and (3) swath of anomalously low intensity pixels on starboard side due to faulty beam pattern.

The third artifact is the swath of low-intensity pixels parallel to and located approximately 3 km athwartships of the ship's track on the starboard side. This diminution of intensity is caused by an irregularity in the beam pattern, which the angle varying gain (AVG) has not been able to correct. The AVG is designed to correct for beam pattern irregularities and the variation of backscatter intensity due to change in the angle of incidence over a uniform bottom. However, system problems or significant differences between the bottom being surveyed and the bottom from which the AVG was designed can result in severe image degradation.

In our method, a parameter set is calculated for both port and starboard sides, consisting of an average pixel intensity (Figure 4a) for strips of the images representing strips of the bottom which would be subtended by 0.25° bins of the beam athwartships. These strips parallel the track of the ship and span the range of angles from nadir to the least depression angle. These averages should be taken along a representative portion of a mosaic so that variations in IV due to local geologic variations will cancel. The transformation is accomplished by multiplying every pixel in the image located at angle increment i by a factor P_i where

$$P_i = \bar{IV} / IV_i \quad (1)$$

and \bar{IV} is the average intensity value of the entire image under consideration and IV_i is the average intensity at the angle $i \times 0.25^\circ$. Comparison of Figures 3 and 4b shows the result of this transformation. The minimization of the three artifacts is evident, while the true geologic features are minimally affected. We refer to this correction as "background subtraction" as it is largely a correction based upon removal of the average image background.

Our method assumes that variations in pixel intensity due to geologic variability will be randomly distributed relative to the track of the ship and thus add destructively in the average as long as the data set is sufficiently large. As the parameter sets are normalized by the overall average of the image, the gain of the filter is independent of the geology or original gain settings. However, spurious features related to the operating system occur at fairly fixed depression angles, adding constructively in the along-track summation, and yielding estimates of systematically induced cross-track errors. Hence the specular reflections (constrained to near-nadir positions), the peg-leg multiples (constrained to a cross-track distance proportional to twice the fish depth), the surface multiples (constrained to a depression angle of 30°), and the beam pattern variations (by definition a function of angle and ray path bending) will all contribute significantly to the average cross-track profile.

Fig. 4a. Plot of average pixel intensity versus look angle for the starboard half of the image shown in Fig. 2. Number 1 pixel is nearest the ship track. Note peaks at 1° , 12° , and 60° corresponding to the specular reflection, peg-leg, and surface multiples, respectively, and the trough about 42° corresponding to the null in the beam pattern.

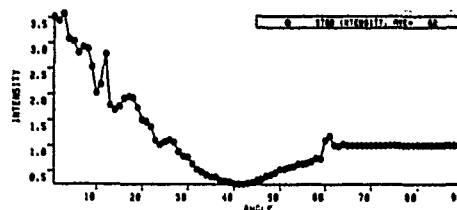
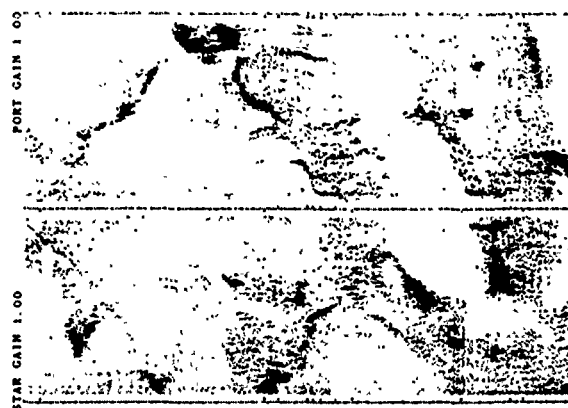


Fig. 4b. Image as shown in Figure 2 after background subtraction. Note removal of specular reflections, surface multiples, and shading problems.



2.2 Geometric Rectifications: Layover

Layover refers to pixels which have been placed at the incorrect cross-track distance. Reflected side scan data from each "ping" (outgoing pulse) are acquired sequentially in time, i.e., linear in slant range. To convert this cross-track "slant range" image to a plan perspective without a priori knowledge of the bottom topography requires that one assume that the bottom is flat across-track. The cross-track positions of the pixels are estimated from knowledge of the travel time, approximate sound velocity, and nadir depth. Violation of this flat bottom assumption, as often happens, will result in topographic features being incorrectly positioned, or "laid over" (Figure 5). Because of this flat bottom assumption, reflections from points A and B in Figure 5, representing off-nadir troughs and peaks, are erroneously rotated along arcs of radii equal to their slant ranges to points A' and B' on the reference datum and consequently are imaged spuriously. As a result, the inward sloping faces of the trough and cliff have been foreshortened, and their boundaries have been misplaced.

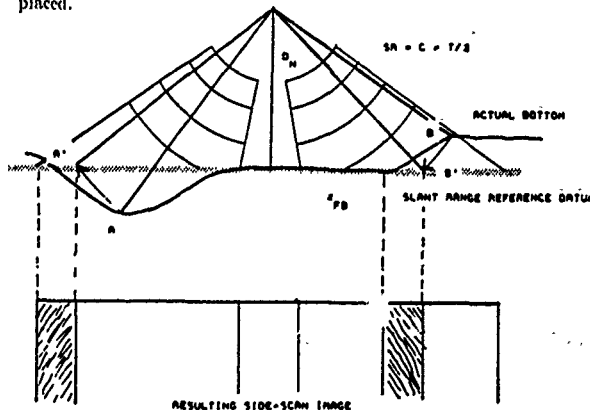


Fig. 5. Slant range correction geometry. Nominal slant range correction assumes that the bottom is flat (reference datum). Topographic deviations from this reference datum will cause features to be imaged incorrectly, as shown by the compression of the inward sloping faces of the trough and cliff.

This displacement is contradictory to our experience from photographic aerial imagery, wherein the taller objects appear to lean outward, away from the viewer. Furthermore, this distortion becomes more significant closer to nadir, again at odds with our experience with aerial imagery in which the distortion decreases with a decrease in parallax view angle. Finally, in mosaics containing parallel tracks with opposite look directions, similar topographic features seen in different tracks will be displaced in opposite directions, resulting in difficulties in coregistration, not to mention any quantitative analysis.

Under the flat bottom assumption, side scan pixels are placed at a cross-track distance X equal to $(SR^2 + TA^2)^{1/2}$ where SR is the slant range (sound velocity * arrival time/2) and TA is the towfish altitude (reference datum - fish depth). While the cross-track position may be wrong, the slant range is correct. Our correction of this pixel position error is accomplished by interpolating a smoothed bathymetric profile at 5 m intervals, and determining the side scan pixel for which the slant range equals that of the nearest bathymetry point X/Z . The results of this transformation are shown in Figures 6, 7, and 8. Note how the axis of the valley, which appears curved in the raw image (Fig. 6) has straightened after processing (Fig. 7) to follow the actual morphology, as given by the bathymetry (Fig. 8). An obvious benefit of this rectification is that features are correctly placed on the image for interpretation and survey targeting. A more subtle benefit is the removal from the image of spurious cross-track compressions and rarefactions due to topography which might otherwise be interpreted by both man and the computer algorithms as variations in geologic character. Furthermore, bottom detect errors, caused by side swipe detection of off nadir bathymetry, and their resulting effects on the side scan image can be explicitly corrected but for small zones of ambiguity in the placement of side scan pixels. These zones of ambiguity occur when multiple bathymetric points yield identical slant ranges. Our heuristic solution to this problem divides the amplitude of the ambiguously located side scan pixel by the number of bathymetry points with that slant range and allocates the quotient to each point.

The only assumption required by our layover correction method is that the side scan data be in the correct sequence athwartships, i.e., that there be no reflector at a cross-track distance X and elevation above the reference datum Z such that the travel time associated with it would be less than that for any reflector located at some $X < X$. This "correct sequence" assumption is a reasonable assumption for most geometries and bottom types, considering the relatively high altitude (height above bottom), typically greater than 20% of the swath width, at which SeaMARC II is deployed.



Fig. 6. Unprocessed side scan image corrected for slant range according to the "flat bottom" assumption.



Fig. 7. SeaMARC II image after layover correction, as described in the text.

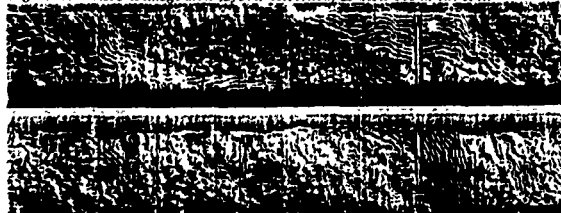


Fig. 8. SeaMARC II bathymetry for the image shown in Figure 6.

3. FEATURE EXTRACTION

3.1 The Feature Vector

The chief purposes of remote sensing imagery are detection and discrimination. Although human observers are clearly capable of interpreting such imagery, the results are subjective, not necessarily repeatable, and often more indicative of the interpreter than the object. Statistical analysis of image texture provides an objective and repeatable means of identifying, distinguishing, and labeling surface types.

3.2 Texture

Texture is an innate property of all objects, which characterizes the closely interwoven relief of the surface. Texture is strongly stationary and independent of illumination. In this paper, we utilize the gray level co-occurrence matrix (GLCM) method of texture analysis [Haralick et al., 1973].

The GLCM method requires the creation of a secondary matrix from which second-order texture statistics are estimated. Specifically, let $F(x,y)$ represent the digital image over a rectangular domain $L_x=1,2,\dots,N_x$, $L_y=1,2,\dots,N_y$, quantized to N_g gray levels. Each GLCM is a square matrix of dimension N_g , whose entries $S(i,j,\theta,d)$ express the number of times there occurs in the image a pixel of intensity i neighbored by a pixel of intensity j in the direction θ , at distance d . The elements of the matrix are normalized by the total number of possible entries R for that direction and lag.

In order to insure that the texture signature of any given nonisotropic texture is not significantly altered by the angle at which it is imaged, these matrices are evaluated for values of θ equal to 0° , 45° , 90° , and 135° . Symmetry considerations allow neglecting the respective supplementary angles.

3.3 GLCM Features

Haralick et al. [1973] suggested 14 features which can be extracted from the GLCMs. Four of these statistics, ASM, CON, ENT, and AIDM, have been shown to be strong estimators of wavelength variations and to be insensitive to variations in either look direction or gain settings [Reed, 1987]. ISO, a feature of our own derivation, shows similar characteristics as the above four features. It is calculated as the sum of the differences of orthogonal GLCMs, and is a measure of the isotropy of the image. The means and ranges for these statistics are evaluated for unit lag. The resulting set of statistics forms a feature vector which describes the texture of that portion of the image over which the features were evaluated and by which that section of the image will be classified. To evaluate the texture of an entire image, the image is usually divided into rectangular cells referred to as "texels." To optimize the potential of this texture routine, these texels should contain at least NG^2 pixels, where NG is the number of gray levels to which the image has been quantized [Pratt, 1978] in order to maximize the possibility of measuring real image texture variations.

3.4 Region Growing -- ReGATA

The texel, the basic unit upon which the analysis is based, is assumed to contain a subset of the image data which possesses only one homogeneous texture. The probability of spatial homogeneity of a randomly placed rectangle increases as the texel size decreases. The texel therefore can neither be shrunk below a minimum size without compromising the textural analysis nor increased drastically without encountering an unacceptable number of mixels.

Rather than divide the image into a priori boxes which may or may not contain homogeneous patterns, we seek, via data-controlled decisions, a subdivision of the image into closed regions of limited spatial heterogeneity. One feature vector will then be used to represent each closed region. By allowing regions to grow to their natural boundaries, the probability of producing "mixels" is strongly reduced. As the regions will in general be larger than with the a priori texels, classification time will be reduced concomitant with the reduction in number of feature vectors.

The following describes the implementation of Region Growing and Texture Analysis (ReGATA) a bottom-up region-growing routine. SeaMARC II data are stored in records, line-interleaved-by-pixel. Two records, containing two pings of 984 port and 984 starboard pixels, are "read in." The data for port and starboard are subdivided into cells containing two lines of 16 pixels each. The mean intensity and variance are calculated for each of these 32-pixel cells. Two tests are then conducted. The first determines if the cell is reasonably homogeneous, in order to separate the cells into the categories of "region" and "boundary." Those cells which do not pass this test of limited heterogeneity are assumed to span two regions and hence are labeled as boundary cells.

The second test is the basis of annexation and region growing. The mean and variance of the cell under consideration is compared to the means and variances of region cells immediately adjacent in the west, northwest, north, and northeast directions. The logic of the comparison is that if two cells are neighbors and possess similar average value, they probably represent a uniform, connected surface in the object plane and hence should be annexed and similarly classified. As a region grows, its mean and variance are recomputed, and the pixels within the annexed cell augment the four GLCMs associated with the region. Upon closure texture vectors are calculated for all regions.

When the routine terminates at the end of the data file, all remaining open regions are closed by the routine and analyzed for texture. The resulting list of region-numbered, feature vectors are submitted to an unsupervised cluster analysis. The region numbers map the regions to the feature vectors in the clusters, and all pixels in all regions with similarly clustered feature vectors are mapped similarly.

3.5 Application of ReGATA

The data set under consideration is of a submarine lava flow located 200 km NW of Vancouver Island, near the Queen Charlotte transform fault and the Tuzo Wilson Knolls (Figure 9). The dark central feature is a lava flow. We assume that the flow is continuous and homogeneous across the ship's track. The symmetric variation in average reflectivity about the ship's track is ascribed to a system AVG ill-adapted for the scattering behavior of the lava flow. Adjacent to the flow is a 100-m mound in the lower right that may be the vent source of the flow. An acoustic shadow, seen here as a patch of low-intensity pixels, occurs on the steep back side of the volcanic mound. The large lava flow apparently crosses the transform fault scarp, which is downdropped to the upper right portion of the image.

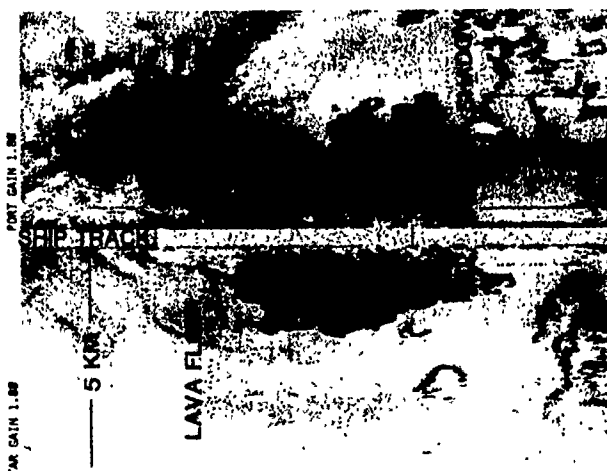


Fig. 9. SeaMARC II side-scan sonar image of a lava flow, located 200 km NW of Vancouver. The linear feature in the top right corner is the Queen Charlotte transform fault scarp, with an associated down drop of several hundred m. Note the 100-m vent from which the flow may have emanated and its acoustic shadow.

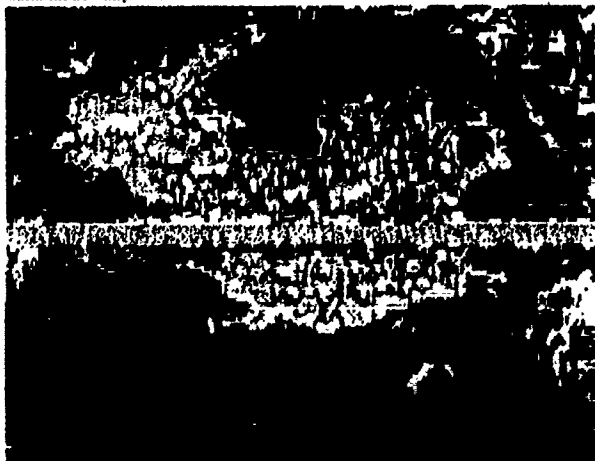


Fig. 10. Three-class ReGATA map of lava flow image shown in Figure 9. Black areas correspond to either "boundary cells" or regions with fewer than five annexed cells. See text for details.

In our application of ReGATA, the side scan image of the lava flow (Figure 9) was re-quantized to 64 gray levels and subdivided into cells containing 32 pixels, for which the average intensities and standard deviations were calculated. A nonparametric, minimally distant means rule was used for region growing, with annexation occurring if the regions differed in their means by three levels or less. All closed regions containing five or more cells were analyzed for texture and hence had associated with them the six-dimensional feature vector. In the cluster analysis which followed, only those regions containing 100 or more cells contributed to the determination of the class centroids. Those regions containing fewer than 100 cells were then placed into the class from which their feature vector was minimally distant.

Subsequent unsupervised classification yields a thematic map (Figure 10) containing three classes, gray, striped, and black, corresponding in the image to regions containing the lava flow and two classes of sediment, respectively. The black area appears to correspond to the less reflective, possibly smoother sediment. The striped areas correspond to visually rougher sediments in the vicinity of the lava flow. In the original side scan image, these sediments appear darker, indicating either greater roughness or greater acoustic impedance. Possibly, the sediments near the flow are mixed with debris from the neighboring lava flow or expressing the roughness of some thinly buried previous flow. The blank areas correspond to either boundary cells or to unannexed regions containing less than five cells. Comparison of feature values for these three classes with those from analysis of synthetic images of known wavelength [Reed, 1987] substantiates this roughness mapping.

4. CONCLUSIONS

Application of the above techniques to a variety of imagery resulted in both superior images for subjective and computer-aided interpretations. Although the use of marine acoustic data, which are strongly influenced by the low speed of sound in water and susceptibility to ray path variations, is inherently more difficult than processing subaerial optical or radar images, we are confident that eventual increased quantification of seafloor data will permit widespread application of various airborne and satellite remote sensing techniques to the imaging and mapping of the seafloor.

REFERENCES

- Blackinton, J. G., D. M. Hussong, and J. Koslos, First results from a combination side scan sonar and seafloor mapping system (SeaMARC II), in *Offshore Technology Conference, OTC 4478* pp. 307-311, 1983.
- Haralick, R. M., K. Shanmugam, and I. Dinstein, Textural features for image classification, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-3, 610-621, Nov., 1973.
- Reed, T. B., IV, Digital image processing and analysis techniques for SeaMARC II side-scan sonar imagery, Ph. D. dissertation, Univ. of Hawaii, Honolulu, 1987.

**LE PROGRAMME SPOT : PLANIFICATION A MOYEN
ET LONG TERME
SUR LA BASE DES BESOINS DU MARCHE**

***SPOT PROGRAMME STATUS: MEDIUM TERM
AND LONG TERM PLANNING
ON THE BASIS OF MARKET REQUIREMENTS***

**Gérard BRACHET
Président Directeur Général, SPOT IMAGE
TOULOUSE, FRANCE**

Dans la conception initiale du programme SPOT aussi bien que dans son évolution à long terme, la ligne directrice principale est la continuité du service rendu aux utilisateurs. Cette priorité donnée à la continuité n'exclut pas l'introduction par étapes d'améliorations diverses destinées à mieux servir les marchés établis ou à répondre à des besoins nouveaux. C'est ainsi qu'au delà des trois premiers satellites SPOT 1, SPOT 2 et SPOT 3, essentiellement identiques, le satellite SPOT 4 incorporera de nombreuses améliorations (durée de vie accrue, bande spectrale dans le moyen infra-rouge, corégistration des données multispectrales à 20 m de résolution et des données monospectrales à 10 m de résolution). Son lancement est prévu pour 1993-94 compte tenu du très bon fonctionnement en orbite du satellite SPOT 1, que l'on espère voir confirmé pour SPOT 2 et SPOT 3.

Au delà de SPOT 4, une nouvelle génération de satellites d'observation optique à haute résolution est à l'étude, comportant un instrument capable de fournir des couples d'images stéréoscopiques le long de l'orbite (donc séparées de seulement quelques minutes) avec une résolution au sol de 5 m. Cette étude est réalisée dans le cadre d'une possible coopération franco-américaine pour cette nouvelle génération de satellites de télédétection pour la deuxième partie de la décennie 90-2000.

EVALUATION OF SPOT HRV IMAGE DATA RECEIVED IN JAPAN

*N.Fujimoto, *Y.Takahashi, *T.Moriyama,
*M.Shimada, *H.Wakabayashi,
**Y.Nakatani,
*** S.Obayashi

* National Space Development Agency of Japan (NASDA)
** Earth Observation Center(EOC)
*** Remote sensing technology center of Japan
Science University of Tokyo

ABSTRACT:

Earth Observation Center(EOC/NASDA) has received and corrected SPOT HRV data from Oct.1988. This paper describes the early results of SPOT HRV evaluation efforts on geometric accuracy, SN ratio, and MTF characteristics. Land cover classification accuracy have also been studied in the application field.

1.INTRODUCTION

NASDA has received SPOT HRV data and has produced radiometric and geometric correction products since Oct. 1988.

The purpose of this study is to evaluate SPOT products and its processing systems in order to improve quality of the products and performance of the systems.

This paper reports some early analytical results for HRV image data received at Earth Observation Center in Japan.

These results will contribute to the development of the Japanese next generation satellites and sensors.

This paper consists of evaluation of radiometric characteristics, geometric correction accuracy, and landcover classification accuracy.

The evaluation of geometric accuracy about HRV data, processed at EOC, is the first topic of this paper. We picked up approximately 30 scenes and evaluated geometric accuracy using GCPs. Cross track and along track direction errors were evaluated concerning each HRV sensor system.

The next topic is on signal to noise ratio of HRV data. The SN ratio was calculated using the software which was developed for MOS-1 verification program.

We also calculated PSF (Point Spread Function), and MTF (Modulation Transfer Function) in some cases to evaluate spacial resolution of HRV.

The last topic is concerning the application field, land cover classification and its accuracy. The result of this study includes joint study between NASDA & Science University of Tokyo.

2.GEOMETRIC ACCURACY

We have selected 28 radiometrically and geometrically corrected HRV scenes and calculated geometric error using 10 to 36 GCPs for each scenes. Table 1 indicates the results. Fig.1 shows the errors of along and cross track directions of each scenes, on the other hand Fig. 2 shows the error changes according to the data acquisition date.

From these data, following results are obtained.

- 1) Along track error has some offset element, (average of the error is -356m) and the magnitude of the errors depends upon the data acquisition date. On the other hand cross track direction error shows random change.
- 2) The same scene observed on the same date with different sensor (XS-1 and PA-1, or XS-2 and PA-2) has almost same error.
- 3) The error of (331,332-278, 88.10.14) indicates the relation between XS-1 and XS-2, and the error of (331-279,332-275,278, 88.6.6) also indicates the correlation between XS-1 and PA-2, these scenes are simultaneously observed.
- 4) Fig.3 illustrates that the mirror angle has some correlation with geometric errors.

We further examined magnification ratio, aspect ratio, rotation angle, and skew angle.

These results are shown in Table 2. Magnification and aspect ratio are small enough so that it will conclude that SPOT orbit estimation accuracy is good about satellite altitude and attitude of the pitch axis.

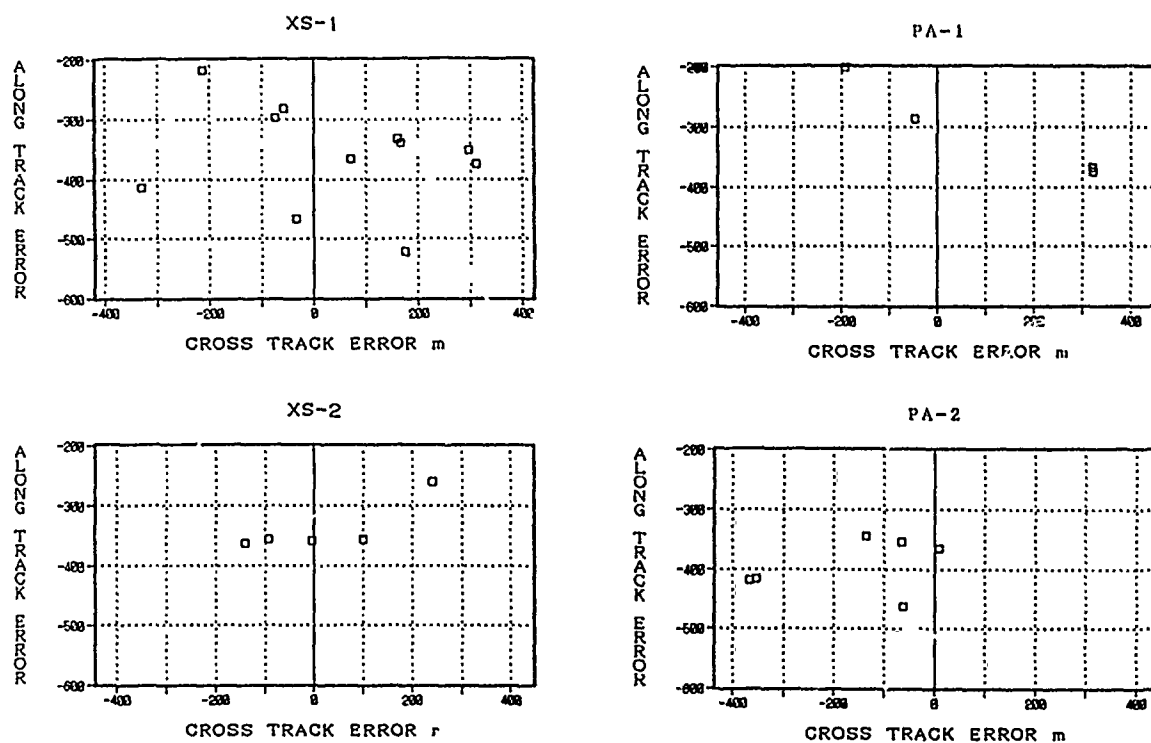


FIG. 1 GEOMETRIC ERROR OF HRV EVALUATED WITH GCPs.

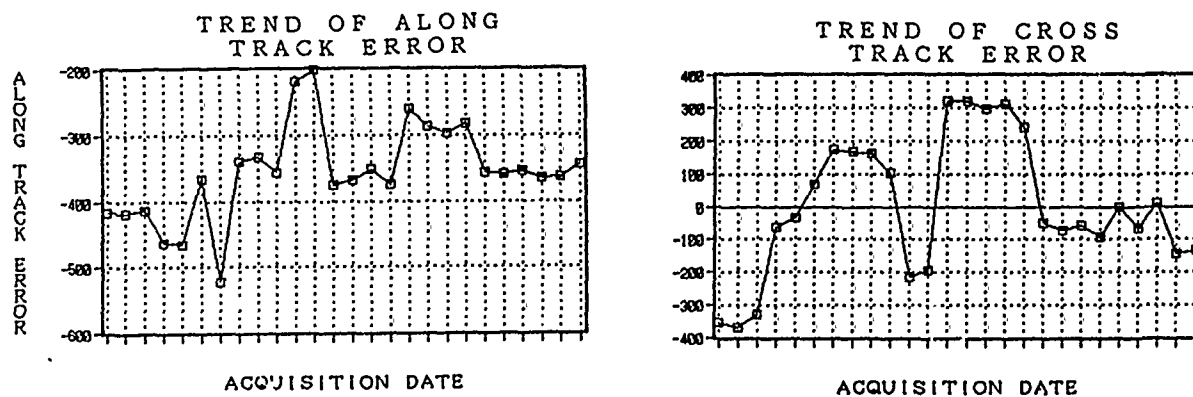


FIG. 2 TREND OF GEOMETRIC ERROR ACCORDING TO DATA ACQUISITION DATE.

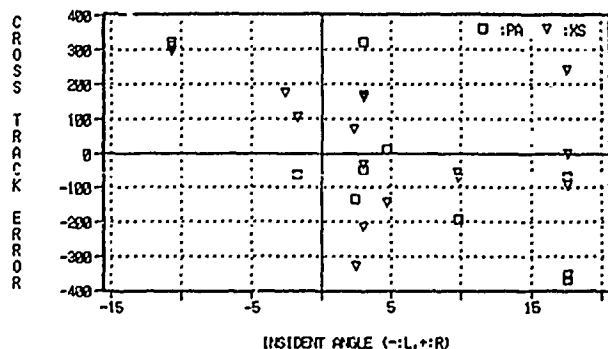


FIG. 3 CORRELATION BETWEEN INCIDENT ANGLE AND CROSS TRACK DIRECTION ERROR.

3. SN RATIO OF HRV

SN ratio were calculated in the 7 positions as follows.

- 1: the mouth of ARAKAWA river
 - 2: urban area of Tokyo
 - 3: forest (Kanto area)
 - 4: the mouth of ARAKAWA river (PA scene)
 - 5: reclaimed land of Tokyo bay
 - 6: the mouth of ARAKAWA river (other scene)
 - 7: the mouth of ARAKAWA river (other scene)
- the results are indicated in the Table 3. and Fig.4.

Each SN ratio was calculated in 32 line by 32 pixel area, and hanging window was used for Fourier Transformation.

The values are about 15dB to 21 dB, and these are not so different from MOS-1 MESSR verification result in the EOC.

Fig.4 shows the characteristics of SN ratio.

1) SN ratio degraded along the spectral band, PA band has particularly low S N ratio in this study.

2) Water area (mouth of ARAKAWA river) shows almost same pattern.

3) In urban area, SN ratio is low in every spectral band. Because urban area includes many bright artifacts and then noisy in frequency domain.

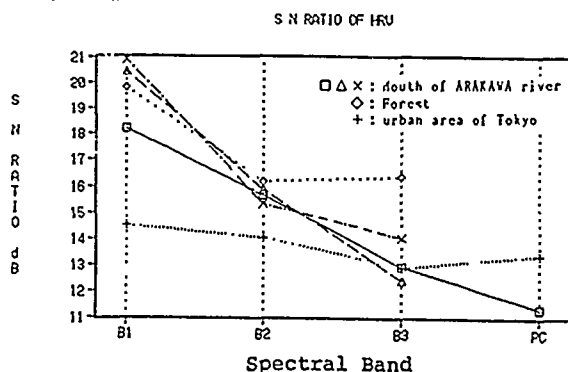


Fig. 4 S N RATIO OF HRV

4. MTF

MTF was calculated by following method.

Select a coast line area from the examined image, and get edge profile by edge tracing, then approximate the profile using spline function. Differentiate this function to get PSF and MTF is calculated from PSF by Fourier Transformation.

These results are illustrated in Fig.5.

The MTFs were good for all other spectral bands. As our verification software idealized some condition in calculation, MTF shows rather ideal curve, that indicates HRV resolution satisfied the specification.

5. CLASSIFICATION

The author also checked the classification result of HRV image data (level 2) using the trace and division accuracy. The trace accuracy means geometric probability. And the division accuracy means geometric probability based on truth data. The classification was done using supervised maximum likelihood classifier. The HRV image data was acquired October 14, 1988 at EOC (Earth Observation Center of NASDA). The study area is Oh-shima, famous volcanic island, the size is 9km east to west and 15km south to north. And this island locates from longitude 139 20'E to longitude 139 27 E, from latitude 34 41'N to latitude 34 48'N.

The table 4 shows the classification accuracy of each category. In the table, the vertical axis is categories of the digital map information got by ground truth. And the horizontal axis is categories for HRV analysis. The situation of error distribution becomes clear from this table.

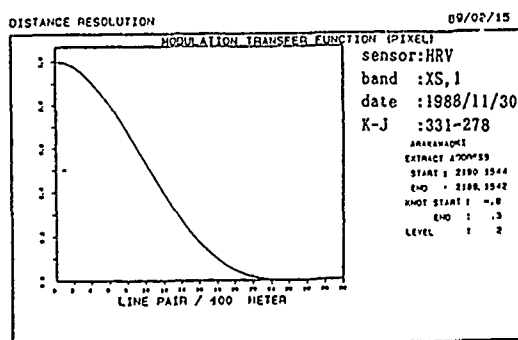
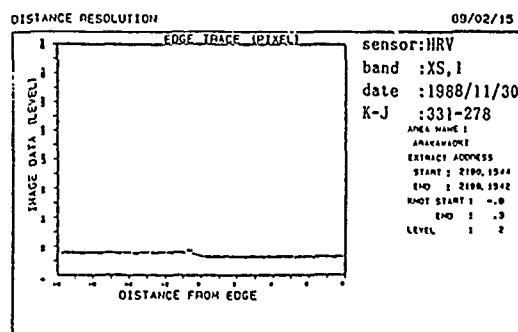


FIG. 5 EDGE PROFILE AND MTF OF HRV IMAGE.

Therefore, it becomes clear that the land cover classification analysis needs other spectral bands.

6. CONCLUSION

HRV data received in Japan was investigated. Concerning to geometric accuracy, alongtrack error will be improved if we consider the offset element. On the other hand we could not find out the characteristics of crosstrack error. Hence SPOT roll axis attitude and its estimation is unstable.

As a conclusion from the S/N analysis water and forest area image data quality is good in XS band. For PA band high resolution will influence to the degradation of SN ratio.

Classification accuracy in this study is not so good. But the values of accuracy were calculated based on 1/25000 classification map data, thus more precise truth data will lead more good result.

However, land classification with maximum likelihood is not so suitable for 2 to 3 spectral bands sensor as HRV.

This study performed on the data received in Japan in early period. Constant verification should also continue to improve the quality of products.

ACKNOWLEDGMENT:

The authors would like to thank HITACH staff for supporting to calculate geometric accuracy.

REFERENCE:

- 1) Analysis and evaluate on LANDSAT, SPOT MOS-1 data, 1989/3/31 S. Obayashi et al Joint study report between NASDA and Science University of Tokyo.
- 2) MESSR IMAGE QUALITY, P. Henry, G. Begin 1989.2 Proceedings of 3rd MOS-1 verification Program.

Table 1 Geometric Accuracy of HRV

| K-J | Mode | Date | Mean | | Standard Deviation | | INCIDENT ANGLE |
|---------|------|----------|--------|--------|--------------------|-------|----------------|
| | | | Line | Pixel | Line | Pixel | |
| 332-275 | P 2 | 88.6.6 | -417.0 | -353.8 | 12.8 | 14.1 | L 2.2 |
| 332-275 | P 2 | 88.8.23 | -462.9 | -63.4 | 14.4 | 12.2 | L 1.6 |
| 332-275 | P 2 | 89.1.11 | -355.7 | -65.2 | 14.5 | 78.9 | R 17.5 |
| 332-275 | X 2 | 89.1.11 | -356.3 | -92.9 | 15.3 | 77.5 | R 17.6 |
| 332-278 | P 2 | 88.6.6 | -419.0 | -368.0 | 20.0 | 14.3 | L 2.2 |
| 332-278 | X 2 | 88.10.14 | -356.3 | 101.3 | 26.3 | 21.4 | L 1.7 |
| 332-278 | X 2 | 88.11.25 | -260.3 | 240.6 | 25.5 | 12.7 | R 12.8 |
| 332-278 | P 2 | 89.1.11 | -365.4 | 9.5 | 17.6 | 15.0 | R 17.5 |
| 332-278 | X 2 | 89.1.11 | -359.0 | -2.7 | 17.2 | 12.3 | R 17.6 |
| 332-278 | P 2 | 89.1.21 | -345.2 | -135.0 | 16.2 | 14.9 | R 4.6 |
| 332-278 | X 2 | 89.1.21 | -363.3 | -111.7 | 18.5 | 17.1 | R 4.7 |
| 331-279 | X 1 | 88.6.6 | -414.6 | -328.7 | 11.6 | 22.0 | R 2.4 |
| 331-279 | X 1 | 88.8.23 | -465.9 | -33.6 | 15.2 | 18.2 | R 3.0 |
| 331-279 | X 1 | 88.10.14 | -333.3 | 160.8 | 13.0 | 20.0 | R 3.0 |
| 331-279 | P 1 | 88.11.19 | -375.4 | 321.3 | 15.3 | 36.3 | L 10.7 |
| 331-279 | X 1 | 88.11.19 | -352.5 | 296.3 | 13.6 | 22.6 | L 10.7 |
| 331-279 | P 1 | 88.11.30 | -286.7 | -48.8 | 13.4 | 30.0 | R 9.7 |
| 331-279 | X 1 | 88.11.30 | -296.7 | -73.8 | 11.8 | 18.4 | R 9.8 |
| 331-278 | P 1 | 88.11.19 | -368.9 | 319.4 | 28.3 | 9.7 | L 10.7 |
| 331-278 | X 1 | 88.11.19 | -375.6 | 312.1 | 51.2 | 23.2 | L 10.7 |
| 331-278 | P 1 | 88.11.9 | -201.7 | -193.2 | 22.8 | 9.0 | R 3.0 |
| 331-278 | X 1 | 88.11.9 | -219.1 | -213.8 | 24.4 | 12.6 | R 3.0 |
| 331-278 | X 1 | 88.10.14 | -339.6 | 166.3 | 23.9 | 11.8 | R 3.0 |
| 331-278 | X 1 | 88.11.30 | -282.1 | -57.8 | 24.1 | 13.5 | R 9.7 |
| 312-284 | X 1 | 88.10.12 | -522.3 | 173.5 | 26.5 | 18.5 | L 2.6 |
| 313-284 | X 1 | 88.10.7 | -367.0 | 70.3 | 39.8 | 17.1 | R 2.3 |

Table 2 Geometric distortion of HRV data

| | | Date | Magnification | | Aspect | | Rotation Skew | |
|---------|-----|----------|---------------|--------|---------|---------|---------------|-------|
| | | | Line | Pixel | Line | Pixel | (deg) | (deg) |
| 331-278 | P 1 | 88.11.19 | 1.0002 | 1.0001 | -0.0077 | -0.0325 | | |
| 331-278 | X 1 | 88.11.19 | 1.0004 | 0.9991 | -0.0036 | -0.0630 | | |
| 331-278 | P 1 | 88.11.9 | 0.9999 | 1.0001 | -0.0124 | 0.0184 | | |
| 331-278 | X 1 | 88.11.9 | 1.0002 | 0.9994 | -0.0065 | 0.0970 | | |
| 331-278 | X 1 | 88.10.14 | 1.0000 | 0.9995 | -0.0196 | 0.0221 | | |
| 331-278 | X 1 | 88.11.30 | 1.0002 | 0.9995 | -0.0142 | 0.0053 | | |
| 312-284 | X 1 | 88.10.12 | 0.9993 | 0.9999 | -0.0053 | -0.0088 | | |
| 313-284 | X 1 | 88.10.7 | 0.9996 | 0.9999 | -0.0067 | -0.0010 | | |

Table 3. S N R of SPOT HRV data

| | | | | | | NOISE | | SIGNAL | | S/N (dB) | *AVERAGE |
|---|----------|---------|----|----|--|----------|----------|--------|--------|----------|----------|
| | | | | | | Line | Pixel | Line | Pixel | | |
| 1 | 88.11.30 | 331-278 | XS | B1 | | 9.54E+03 | 4.13E+07 | 18.184 | 15.050 | | |
| | | | | | | 5.69E+03 | 7.71E+06 | 15.657 | 8.270 | | |
| | | | | | | 2.80E+03 | 1.08E+06 | 12.935 | 3.080 | | |
| 2 | | | XS | B1 | | 1.06E+05 | 8.47E+07 | 14.509 | 27.150 | | |
| | | | | | | 5.79E+04 | 3.72E+07 | 14.040 | 18.100 | | |
| | | | | | | 8.69E+04 | 3.25E+07 | 12.866 | 16.860 | | |
| 3 | | | XS | B1 | | 5.28E+03 | 4.83E+07 | 19.805 | 20.710 | | |
| | | | | | | 1.09E+04 | 1.88E+07 | 16.179 | 13.130 | | |
| | | | | | | 3.15E+04 | 5.87E+07 | 16.349 | 23.260 | | |
| 4 | | | PC | B1 | | 1.56E+05 | 7.29E+07 | 13.344 | 25.730 | | |
| | | | | | | 1.40E+05 | 2.60E+07 | 11.345 | 15.200 | | |
| | | | | | | 5.31E+03 | 6.51E+07 | 20.440 | 23.830 | | |
| 6 | 88.10.14 | 331-278 | XS | B2 | | 9.75E+03 | 1.45E+07 | 15.855 | 11.240 | | |
| | | | | | | 8.26E+03 | 2.48E+06 | 12.387 | 4.960 | | |
| | | | | | | 2.76E+03 | 4.18E+07 | 20.905 | 19.140 | | |
| 7 | 88.11.30 | 331-279 | XS | B2 | | 7.36E+03 | 8.54E+06 | 15.322 | 8.540 | | |
| | | | | | | 1.67E+03 | 1.07E+06 | 14.037 | 2.050 | | |

*AVERAGE of pixel value

S/N EVALUATION AREA SIZE 32*32
WINDOW FUNCTION HANNING
SIGNAL AREA SIZE 15

Table 4 Classification Accuracy

| | FARM FIELD | CONIFER | BROAD LEAVED | ARTIFICIALITY | ROCK FIELD | BARRE FIELD | GRASS FIELD | LAWN FIELD | SEA | CLOUD | SHADOW OF CLOUD | CLASTICITY | TRACE ACCURACY |
|-----------------|------------|---------|--------------|---------------|------------|-------------|-------------|------------|--------|-------|-----------------|------------|----------------|
| FARM FIELD | 23.87 | 11.86 | 59.23 | 2.56 | 0.01 | 1.91 | 0.01 | 0.41 | 0.00 | 0.00 | 0.00 | 11.09 | 23.87 |
| CONIFER | 19.32 | 21.06 | 56.79 | 0.97 | 0.00 | 1.30 | 0.00 | 0.54 | 0.00 | 0.00 | 0.00 | 0.13 | 21.06 |
| BROAD LEAVED | 9.06 | 5.28 | 78.09 | 1.94 | 0.15 | 4.16 | 0.38 | 0.65 | 0.00 | 0.00 | 0.00 | 13.65 | 78.09 |
| ARTIFICIALITY | 4.27 | 2.13 | 41.88 | 28.69 | 1.11 | 20.83 | 0.21 | 0.86 | 0.00 | 0.00 | 0.00 | 1.61 | 28.69 |
| ROCK FIELD | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| BARRE FIELD | 1.51 | 3.55 | 35.51 | 3.38 | 9.35 | 45.95 | 0.36 | 0.10 | 0.29 | 0.00 | 0.00 | 2.31 | 45.95 |
| GRASS FIELD | 18.17 | 1.18 | 48.22 | 15.79 | 0.15 | 8.77 | 1.28 | 3.13 | 0.00 | 0.00 | 0.00 | 2.71 | 1.28 |
| LAWN FIELD | 2.39 | 2.59 | 76.43 | 7.23 | 0.32 | 8.51 | 0.22 | 2.31 | 0.00 | 0.00 | 0.00 | 1.84 | 2.31 |
| SEA | 0.00 | 0.06 | 0.07 | 0.09 | 0.00 | 0.85 | 0.00 | 0.00 | 98.93 | 0.00 | 0.00 | 46.99 | 98.93 |
| CLOUD | 2.31 | 3.20 | 55.65 | 1.52 | 7.93 | 28.22 | 0.51 | 0.62 | 0.00 | 0.00 | 0.00 | 18.10 | 0.00 |
| SHADOW OF CLOUD | 5.55 | 2.75 | 55.50 | 1.59 | 12.08 | 19.26 | 0.00 | 3.26 | 0.00 | 0.00 | 0.00 | 1.80 | 0.00 |
| CLASTICITY | 5.09 | 3.04 | 32.35 | 2.99 | 1.92 | 8.46 | 0.20 | 0.46 | 46.49 | 0.00 | 0.00 | | |
| TRACE ACCURACY | 52.05 | 3.01 | 31.50 | 23.57 | 0.00 | 12.54 | 17.29 | 9.24 | 100.00 | 0.00 | 0.00 | | |

THREE-DIMENSIONAL ANALYSIS OF SPOT HRV DATA

Masatoshi MORI and Keinosuke GOTOH*

Department of Management Engineering,
Kinki University, Iizuka 820, Japan

*Department of Civil Engineering,
Nagasaki University, Nagasaki 852, Japan

Abstract

A statistical analysis is performed in the three-dimensional (3-D) histogram space of SPOT HRV data. The distribution of vectors (pixels) in the histogram space of HRV data may be different from other satellite sensors such as Landsat TM because of its higher spatial resolution. The SPOT HRV system has the three channels of sensors, which generate a 3-D histogram space in feature. A part of the present objective is to examine the distribution of vectors in this 3-D histogram space. First, we analyze the two-dimensional (2-D) histogram of HRV data, which gives the contour map. Some clearly isolated clusters could be detected in this histogram map. Each cluster contains several contents, which represent land cover features. Then, we develop the algorithm of calculating the 3-D histogram of HRV data. The result of estimating the distribution is given in a histogram table. Also, 3-D peaks in the histogram are detected, and given in a table. Using these tables, the 3-D view system of the 3-D histogram is developed. The 3-D surface with an equalled-value is displayed on a graphic monitor.

Keywords: SPOT, histogram space, 3-D analysis, 3-D peaks

1 Introduction

The objective of the present investigation is to examine the distribution of vectors (pixels) in the three-dimensional (3-D) histogram space of SPOT (Système Probatoire d'Observation de la Terre) HRV (Haute Resolution Visible) data and to compare the peaks in this distribution with land cover features. The spatial resolution (20m) of HRV data is so superior to the previous ones acquired by civilian remote sensing satellites, which approaches the level provided by aerial photography. Furthermore, another feature of the SPOT satellite system is the extended function of viewing up to 27 degrees off-nadir, which gives a stereoscopic image from two different orbits.

First, we analyze one- and two-dimensional (1-D, 2-D) histograms of HRV data to understand basic statistical characteristics of data. For the reason, we select the scene (512x512 pixels) as shown in Fig.1, which contains typical land features such as water,

pond, river, residential, commercial, grass, etc. This scene was obtained by SPOT on October 20, 1986 (ID 313- 283). After calculating 1-D histogram of HRV data and cross-correlations between each pair of sensors, we analyze 2-D histogram of the same data, which is described in Sec.II.

Next, we try to examine distributions of vectors in the histogram space. In HRV data, the histogram space is 3-D. Each pixel in the image as shown in Fig.1 is corresponding to a certain cell in the histogram space, which has three peculiar values obtained from three sensors. In a flat area such as water, many pixels of it belong to one cell. So we calculate a distribution table of the 3-D histogram of Fig.1. Also, 3-D peaks in the histogram space are detected, which may represent land cover features. Utilizing the histogram table, we construct a 3-D view system of the 3-D histogram of HRV data. Setting an equalled-value of counts, the surface in the 3-D histogram is displayed on a graphic monitor.

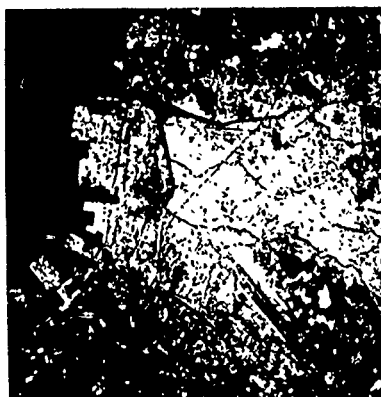


Fig.1. Present study area.

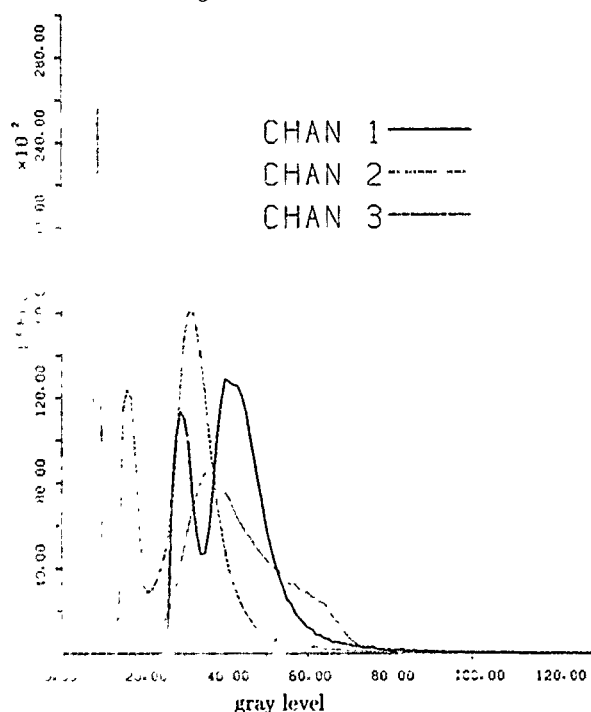
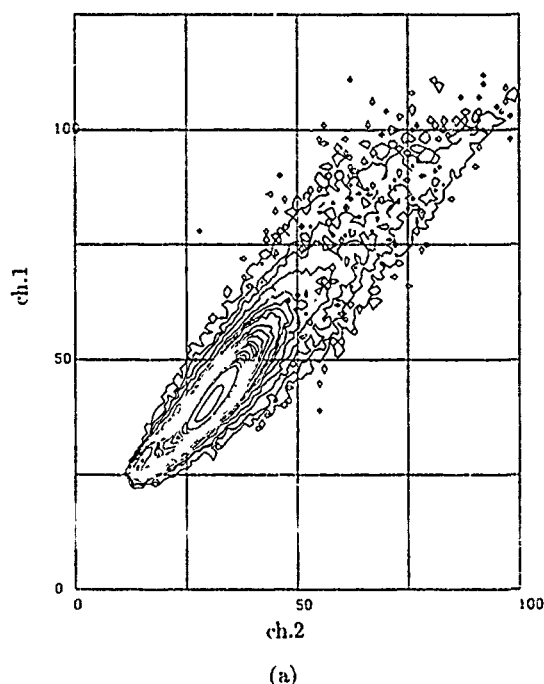


Fig.2. One-dimensional histogram of HRV data.

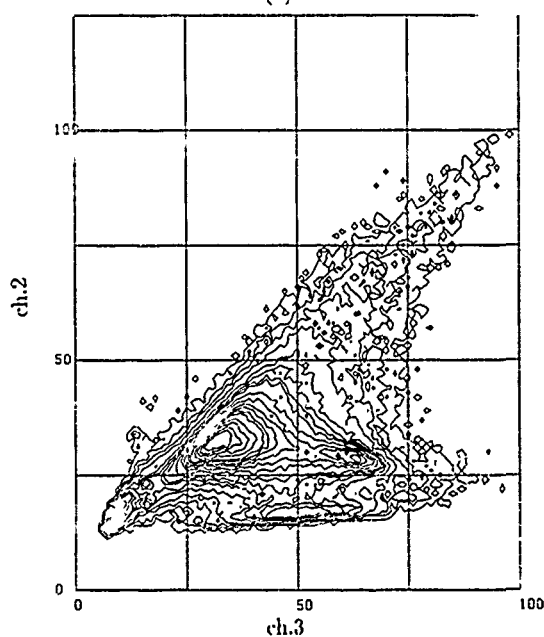
II One- and two-dimensional analysis of SPOT HRV data

First, we show the 1-D histograms of Fig.1 in Fig.2. This Figure shows a plot of gray level versus frequency of pixels. Contrary to our expectation, the effective dynamic range of HRV sensors is not so wide, compared with previous satellite sensors such as Landsat TM. In this scene, the values of the effective dynamic range are less than 100.

Second, we calculate 2-D histograms of HRV data so as to make clear the relation between two channel sensors. Figure 3 shows the 2-D histograms of HRV data in a contour map. In Fig.3(a) we show the relation between channels 1 and 2, and in Fig.3(b) the



(a)



(b)

Fig.3. Two-dimensional histogram of HRV data.

Table I. Cross-correlation.

| | ch.1 | ch.2 | ch.3 |
|------|-------|-------|-------|
| ch.1 | 1.000 | - | - |
| ch.2 | 0.949 | 1.000 | - |
| ch.3 | 0.508 | 0.596 | 1.000 |

relation between channels 2 and 3. Because of strong correlation between sensors as shown in Fig.3(a), this histogram is so trivial that we can get not so much information from it. On the other hand, Figure 3(b) shows quite different distributions from Fig.3(a). The difference between them originates from the difference between their cross-correlation. So we calculate the cross-correlations between each pair of sensors (Chavez, 1984), and show the results in Table I. The value of cross-correlation between channels 1 and 2 is near 1.0, which means that these two sensors have almost same characteristics. This result is corresponding to the distribution of pixels in Fig.3(a). The values of cross-correlation between channels 1 and 3, and 2 and 3 in Table I are less than 1.0, so these characteristics of each pair are rather independent each other. Therefore, several clusters appear in the distribution of pixels in Fig.3(b). Each isolated cluster contains several contents, which represent land features.

By extracting typical training-fields from Fig.1, and projecting them to Fig.3(b), we can recognize

contents of each cluster (Mori, 1988). The isolated cluster with a sharp peak, lower left in Fig.3(b), is corresponding to water area; sea, pond and river. The long and thin cluster, lower middle, is corresponding to wood. The large triangular cluster, upper middle, is corresponding to several contents; residential, commercial and other urban area. The small thin cluster, right, is corresponding to the several types of grass. Comparing the distribution in Fig.3(b) with the plot in Fig.2, we can recognize contents of sharp peaks in lower sides in the histogram of Fig.2. The sharp peak of the plot of channel 3 contains only water area, so we can clearly divide the image of Fig.1 into two parts; water and land area. On the other hand, although the plot of channel 2 is similar to one of channel 3, the sharp peak contains water area and wood. In the plot of channel 1, the sharp peak contains the same contents as one of channel 2 due to strong correlation between channels 1 and 2.

In the 2-D histogram, however, it is not sufficient to recognize all contents corresponding to real land cover features correctly. Even if we analyze the dis-

Table II. 3-D distribution of pixels.

| Counts | Number of cells |
|--------|-----------------|
| 1 | 12294 |
| 2 | 4077 |
| 3 | 2177 |
| 4 | 1393 |
| 5 | 1045 |
| 6 | 783 |
| 7 | 662 |
| 8 | 609 |
| 9 | 465 |
| 10 | 400 |
| 11 | 351 |
| 12 | 332 |
| 13 | 279 |
| 14 | 277 |
| 15 | 249 |
| 16 | 228 |
| 17 | 212 |
| ... | ... |
| 440 | 1 |
| 618 | 1 |
| 635 | 1 |
| 651 | 1 |
| 705 | 1 |
| 776 | 1 |
| 900 | 1 |
| 937 | 1 |
| 1042 | 1 |
| 1091 | 1 |
| 1369 | 1 |
| 1383 | 1 |
| 1433 | 1 |
| 1853 | 1 |
| 1861 | 1 |
| 1905 | 1 |
| 2595 | 1 |
| 2759 | 1 |
| 2807 | 1 |
| 2965 | 1 |
| 2997 | 1 |
| 3969 | 1 |

Table III. 3-D peaks in the histogram space.

| Peak No | ch.1 | ch.2 | ch.3 | Counts |
|---------|------|------|------|--------|
| 1 | 28 | 15 | 8 | 2965 |
| 2 | 29 | 16 | 8 | 2807 |
| 3 | 29 | 18 | 53 | 86 |
| 4 | 31 | 17 | 9 | 3969 |
| 5 | 32 | 18 | 9 | 1861 |
| 6 | 33 | 19 | 9 | 776 |
| 7 | 38 | 28 | 67 | 90 |
| 8 | 38 | 29 | 62 | 91 |
| 9 | 38 | 29 | 65 | 125 |
| 10 | 38 | 29 | 68 | 100 |
| 11 | 39 | 30 | 60 | 110 |
| 12 | 39 | 30 | 64 | 128 |
| 13 | 40 | 29 | 28 | 85 |
| 14 | 40 | 30 | 27 | 82 |
| 15 | 40 | 30 | 31 | 101 |
| 16 | 40 | 30 | 34 | 90 |
| 17 | 40 | 31 | 58 | 87 |
| 18 | 40 | 31 | 63 | 118 |
| 19 | 41 | 30 | 28 | 98 |
| 20 | 41 | 31 | 36 | 125 |
| 21 | 42 | 31 | 31 | 142 |
| 22 | 42 | 31 | 38 | 113 |
| 23 | 42 | 32 | 42 | 85 |
| 24 | 43 | 32 | 30 | 150 |
| 25 | 43 | 32 | 32 | 168 |
| 26 | 43 | 32 | 35 | 138 |
| 27 | 43 | 32 | 39 | 112 |
| 28 | 44 | 32 | 31 | 171 |
| 29 | 44 | 33 | 32 | 170 |
| 30 | 44 | 33 | 34 | 155 |
| 31 | 45 | 33 | 38 | 128 |
| 32 | 45 | 34 | 33 | 174 |
| 33 | 45 | 34 | 37 | 143 |
| 34 | 45 | 34 | 39 | 119 |
| 35 | 45 | 34 | 41 | 95 |
| 36 | 46 | 35 | 34 | 150 |
| 37 | 47 | 36 | 36 | 118 |
| 38 | 48 | 37 | 36 | 107 |
| 39 | 49 | 36 | 35 | 128 |
| 40 | 50 | 37 | 34 | 83 |
| 41 | 50 | 37 | 37 | 90 |

tribution in Fig.3(b), we cannot take a contribution of channel 1 into account.

III The three-dimensional distribution in the histogram space

We calculate a table of the 3-D distribution of pixels in the histogram space to understand the characteristics of SPOT sensors. As setting the effective dynamic range of sensors as 100, as seen in Fig.2, one million cells are necessary to construct the 3-D distribution. Each pixel in the image belongs to a peculiar cell of them. Most of one million cells, however, are empty, because many pixels belong to a small number of cells. So computer memory for one cell needs two bytes at least, which count numbers of pixels up to 65536. Then, computer memory needs 2 mega bytes for all cells. We show a part of the result of the distribution in Table II. The values in the left column in Table II means a number of pixel counts for a cell, and also the value in right means a number of cells in the 3-D histogram. The value of pixel count, a top row, means that these cells have only one pixel count, and almost domain among not empty cells. The value 3969 of pixel count, a bottom row, is the maximum one, which cell is found to belong to sea compared with the 2-D histogram in Fig.3(b). The total number of not empty cells is 29099, which is 2.91% for all cells.

Next, we search peaks in the 3-D histogram utilizing the above-mentioned table. A peak in the 3-D histogram is detected by scanning the three directions in the histogram space, which is defined to have a local maximum value of counts of pixels. This peak seems to be a core representing a certain land cover feature such as sea, grass, residential etc. We show the result in Table III. The value in a left column means a peak number, next three values mean the gray levels of three channels, and the value in right means the number of counts of pixels of core cells. Here we show a part of the peaks, which counts of pixels are more than 80. The peaks of No.1, 2, 4, 5, 6 in Table III belong to sea in Fig.1, which have a large number of counts because of flat area. We can use these peaks as seeds in clustering scheme such as k-means method.

IV The three-dimensional view system of the histogram space

Utilizing the table of the 3-D histogram in Sec.III, we try to construct the 3-D view system of them. This view system could show a 3-D picture of the 3-D histogram. A part of the result is shown in Fig.4. This Figure shows the equalled-value surface of the 3-D histogram in the space. It is impossible to display

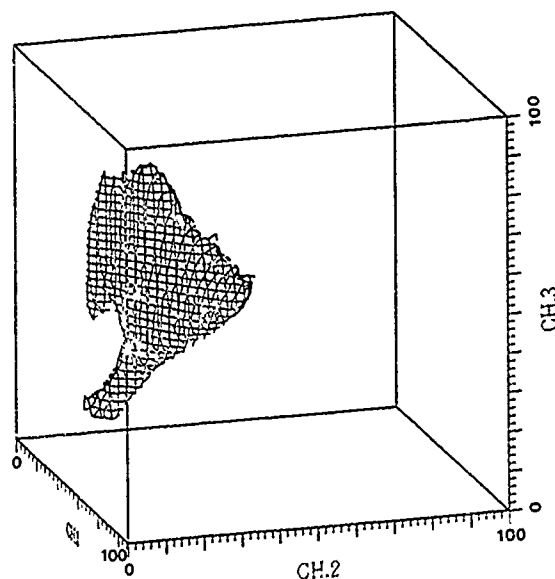


Fig.4. Three dimensional view of the histogram surface.

all cells in one direction of a view point, even if using the 3-D view system. Then, we defined a surface with an equalled-value of the histogram by setting a threshold value of the histogram. In Fig.4, we set the threshold value as 40, and drew the picture by the laser printer. Each value of the axes means the gray level of channels. In this system we can set an any direction of a view point. Decreasing the threshold value, cores of isolated clusters appear one by one. Each of these cores is corresponding to land cover features.

Acknowledgements

We wish to thank the staffs of the computer center of Kyushu University for use of the image processing facility FIVIS.

References

1. Chavez, Jr. P.S. and Berlin, G.L., "Digital Processing of SPOT Simulator and Landsat TM Data for the SP Mountain Region, Arizona", Proceedings of the 1984 SPOT Symposium, pp.56-66, 1984.
2. Mori, M. and Gotoh, K., "Statistical Analysis of SPOT HRV/PA Data", International Archives of Photogrammetry and Remote Sensing, Vol.27, Part B, pp.366-373, 1988.

REMOVAL OF THE TOPOGRAPHIC EFFECT FROM SPOT-1 HRV MULTISPECTRAL IMAGERY IN MOUNTAINOUS TERRAIN

Chongjun YANG - Alain VIDAL - Laboratoire Commun de Télédétection CEMAGREF-ENGREF - B.P. 5095 - 34033 MONTPELLIER
CEDEX 1 - FRANCE

ABSTRACT

Radiance data detected by satellite sensor is generally influenced by undesirable topographic effect. Removing this effect is important in order to optimize information extraction from remotely-sensed data. A new approach as an answer to this necessity is proposed in this paper. The digital numbers in SPOT-1 HRV multispectral imagery are transformed to values of reflectance factors being intrinsic characteristic of the surface material being imaged and invariant to topography. This approach is based on a radiative transfer model that is developed as an approximation of changes in radiance. The values of some unknown parameters of this model were estimated by using multiple linear regression in a small homogeneous test site of 508 pixels, located within the study area. The calculated values of reflectance factors were presented by dimensionless digital numbers ranging from 0 to 255 which corresponds respectively with the values of reflectance factors from 0 to 1. The feasibility of this approach have been tested against the SPOT-1 HRV multispectral images above a canopy of Picaussel Forest in France.

Key-words : SPOT-1 HRV multispectral imagery, Topographic effect, Digital Elevation Model, Reflectance factor, Multiple linear regression.

1 - INTRODUCTION

The topographic effect is defined as the variation in response from inclined surfaces compared to response from a horizontal surface as a function of the orientation of the surface relative to the light source and sensor position. Radiance data detected by remote sensing satellite is generally influenced by the undesirable effect. In areas of high relief, such effect can produce erroneous results when using satellite data for land cover classification and mapping. Many researchers in remote sensing have demonstrated that topographic effect may be quantified and reduced (WOODHAM 1987, and many others).

The spectral reflectance of a surface material is of large interest in remote sensing and satellite digital images are commonly analysed by using the digital numbers for each pixel recorded on a computer-compatible magnetic tape. Thus, An approach as a method of the removal of the topographic effect from the satellite digital images is proposed. That is, the digital numbers are transformed to values

of reflectance factors being intrinsic characteristic of the arbitrary flat surface material being imaged and invariant to topographic.

The layout of the paper is as follows : in section 2, the radiative transfer model is described. An algorithm proposed for determining the values of unknown parameters of the model, an application example and the conclusions are given in section 3.

The presented approach is used within the limits of visible and near-infrared parts of the electromagnetic spectrum and is, in essence, applicable to suitable LANDSAT image data.

2 - MODELISATION

The spectral radiance, coming from the sun, attenuated by the atmosphere, reflected by an Earth's surface, reattenuated by the atmosphere and entering into the sensor of the satellite is influenced to a greater or lesser extent by a large number of effects which may be, in principle, classified as astronomical, geographical, geometrical, atmospheric, meteorological, topographic and biological factors. As an output, the digital numbers, which the users obtain, are modulated by opto-electronic effects.

An idea mathematical form of the transformation from digital numbers to the values of the reflectance factors should contain a full treatment of the above factors. But it is currently not feasible. To date, in remote sensing applications, many simplifications have been made to deal with this complex situation such as assuming isotropic sky radiance distribution and Lambertian surface. In this section, the principal physical phenomena, the mathematic form of transformation, and suitable assumptions will be described. In all the following expressions, the wavelength subscriptions have been suppressed.

The mathematical form of the transformation may be derived from the definition of the BRDF and from the corresponding physical definitions. The detail of the derivation has been omitted here. In this study, the adjacent slope radiances were ignored. The skylight was treated as a uniform hemispherical source, in the same way, the direct sunlight as a well-collimated source (parallel rays), the atmosphere as a series of horizontally homogeneous parallel slabs, and the ground cover as an arbitrary

flat lambertian surface. The term "arbitrary flat surface" was assumed as a mathematically continuous flat surface on a macroscopic scale, having an arbitrary slope and aspect. The altitude-related atmospheric effects were neglected. The spectral radiance at the sensor of satellite L_s ($W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$) consists of that due to the solar radiation reflected by the target (the reflected solar radiation is the sum of the irradiances of the direct sunlight and of the skylight) and reattenuated by the atmosphere, and that (L_r , units $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$) due to backscatter from the direct solar ray including single scattering, multiple scattering and the reflected part by the ground (the environment of the target) into atmosphere (YANG 1989), that is,

$$L_s = \frac{(R_o/R)^2 S_o T_1 \cos I + \pi F \cos^2(\alpha/2)}{\pi} p T_r + L_a \quad (1)$$

where

$$T_1 = e^{-T/\cos S} \quad T_r = e^{-T/\cos V}$$

T_1 and T_r are the total atmospheric transmittance corresponding atmospheric attenuations between the sun and the earth's target, and the earth's target and the sensor, respectively. S (rad) is the solar zenithal angle, V (rad) the zenithal angle of the sensor and T the total normal optical thickness, which equals the sum of the separate optical thicknesses of all the attenuating constituents. I (rad) is the angle between the incident solar ray and the surface normal, which is equal to the solar zenithal angle when the surface is horizontal. Here, α (rad) is the angle of inclination of the local surface (slope), and F the spectral sky scattered radiance ($W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$). S_o is the solar spectral irradiance ($W \cdot m^{-2} \cdot \mu m^{-1}$) outside the atmosphere. p (dimensionless) is the reflectance factor. The effect of the actual Sun-Earth distance R being generally different from its mean value R_o (1 A.U.) is taken into account by the factor $(R_o/R)^2$. The relation between the spectral radiance and digital numbers DN (dimensionless) in the given band in SPOT-1 HRV multispectral imagery is quantified as

$$DN = K \cdot L_k \quad (2)$$

where L_k is the SPOT equivalent spectral radiance (BEGNI 1982), and K the absolute calibration coefficient ($W^{-1} \cdot m^2 \cdot sr \cdot \mu m$). Let S_{oa} , T_{ia} , F_a , p_a , T_{ra} , L_{aa} be the weighted average of S_o , T_1 , F , p , T_r , L_a in a given spectral band, respectively. The reflectance factor can be written as

$$p_a = \frac{(DN/K - L_{aa})\pi}{((R_o/R)^2 S_{oa} T_{ia} \cos I + \pi F_a \cos^2(\alpha/2)) T_{ra}} \quad (3)$$

Based on the above equation, when all the unknown parameters are obtained, the reflectance factors p_a can be calculated.

3 - APPLICATION

3.1. Study area and data sets

A 7.2 km by 3.8 km mountainous area located within the PICAUSSEL Forest in France was selected to test the transformation approach. The latitude and longitude of this area are 42°52'N, 2°25'E,

respectively. This study area is characterized by relief from 550 m to 1160 m above sea level. The steepest slope was 62° and most of the slopes were about 14-20°.

The SPOT-1 HRV image (level 2) for the study area is K44, J246; acquired on 13 August 1986 with a view angle of 3.7°.

Digital Elevation Data (DED) were generated by digitizing the contour lines of the 1:25000 scale topographic map (20m-50m contour interval). The Digital Elevation Model (DEM) of the study area was then created by interpolation of the digital elevation data. The software used was taken from CNES (PROY 1986).

3.2. Computations of the parameter values

The images of real values of slope and aspect of the local surface were derived from the DEM by computing partial derivatives of the local surface of the terrain model. The image of real values of cosine of the incident of solar radiation was generated by using the following equation

$$\cos I = \cos \alpha \cos S + \sin \alpha \sin S \cos(\theta_\alpha - \theta_s) \quad (4)$$

where α and θ_α are the zenithal angle of the local surface normal (or slope) and the azimuthal angle of the local surface normal (or aspect). S and θ_s are the zenithal and azimuthal angles of the solar radiation. The following equation was used to compute the mean-distance correction factor $(R_o/R)^2$.

$$\frac{R_o}{R} = \frac{(1 + e \cos n(J-4))}{(1 - e^2)} \quad (5)$$

where

$$n = 2\pi / 365.25$$

$$e = 0.01673$$

the e is the excentricity of the Earth's orbit, J the Julian day. The value of mean solar irradiance (S_o) for each SPOT band can be obtained by integrating the standard spectral solar irradiance curve over the corresponding wavelength interval. The values of K for each band are directly available from SPOT-1 HRV imagery (CNES 1986).

3.3. Multiple linear regression

Equation (3) can be formally rewritten as

$$DN = A \cos I + B \cos^2(\alpha/2) + C \quad (6)$$

$$\begin{aligned} \text{with } A &= K p_a (R_o/R)^2 S_{oa} T_{ia} T_{ra} / \pi \\ B &= K p_a F_a T_{ra} \\ C &= K L_{aa} \end{aligned}$$

$$\text{and } T_{ia} = e^{-T_a/\cos S} \quad T_{ra} = e^{-T_a/\cos V}$$

For a small homogeneous test site, it was assumed that each point of the test site was identical in the reflectance factor p_a . Because the atmosphere was treated as a series of horizontally homogeneous parallel slabs, the values of the parameters K , $(R_o/R)^2$, S_{oa} , T_{ia} , T_{ra} , L_{aa} for all point of study area could be assumed to be constant. Hence, the parameters A , B , C for all points of the test site were constant. In this study, using $\cos I$ and $\cos^2(\alpha/2)$ as regression variables, the values of A , B , C for each band were estimated by using multiple linear regression over a homogeneous test site of 508 pixels, located within the study area.

3.4. Estimation of the values of the unknown parameters

The methods proposed in section 3.2 were used to compute the values of the parameters $(R/R)^2$, S_{oa} , I , α . The values of parameters K , DN , S , V were directly available from SPOT-1 HRV images. In this study, the values of the total normal optical thickness τ_a , the mean path radiance L_a , the mean diffuse radiance F_a for each band of a SPOT-1 HRV image, needed in equation (3) were evaluated by using the following method. Let p_{ao} be the exact value of the reflectance factor for the test site, τ_{ao} the exact value of the total normal optical thickness. It was assumed that p_{ao} existed between p_{am} and p_{an} ($m < n$), and τ_{ao} existed between τ_{am} and τ_{an} . p_{am} , p_{an} , τ_{am} and τ_{an} were known as the empirical ancillary information. For each couple of p_{ai} , τ_{ai} ($m \leq i, j \leq n$), which vary between p_{am} and p_{an} and between τ_{am} and τ_{an} respectively, using equation (6), one obtained a certain $DN_{i,j}$ (S_{oa} , I , α , $(R/R)^2$, K , S , K were known, $F_a = B/(K p_{ai} \tau_{ai})$) and the value of the point variance for the calculated value $DN_{i,j}$ and the original digital numbers DN . Further on, the total variance for all points of the test site was calculated, respectively to each couple of p_{ai} , τ_{ai} . Thus, a couple of p_{aa} , τ_{aa} ($m \leq A, B \leq n$) which made the total variance minimum was chosen as a couple of the approximate values of the exact values p_{ao} , τ_{ao} . Note that p_{aa} , which was not identified in all the study area was only used to estimate the value of the total normal optical thickness and to compute the value of the mean spectral diffuse radiance F_a ($F_a = B/(K p_{aa} \tau_{aa})$).

Up till now, the values of all the parameters needed in equation (3) have been computed.

3.5. Results and conclusions

A digital elevation data image is shown in figure 1. Dark areas represent low elevation, light areas represent high elevation, light areas represent high. A slope image derived from the DEM and a $\cos I$ image generated using equation (4) are shown in figure 2 and in figure 3, respectively. Figure 4 and figure 5 show a portion of SPOT-1 HRV image (band 1) of Picaussel Forest and a corresponding synthetic reflectance factor image generated using equation (3), respectively.

Visual examination of uncorrected image (Fig. 4) and corrected image (reflectance factor image) (Fig. 5) reveals that the shadows are partially removed.

The method proposed in this paper requires a priori knowledge regarding a dynamical range of the reflectance factor variation of selected homogeneous test site, but not regarding a value of the reflectance factor. The homogeneous test site needs be small enough in study area that the reflectance factor can be considered constant throughout.

In this study, only lambertian reflection was considered. Non-lambertian reflection law will be considered in future research.

ACKNOWLEDGMENTS

The contributions of A. Polo, C. Puech, P. Maurel, H. Piaton and M. Deshayes (LCT CEMAGREF-ENGREF) in this investigation are appreciated.

REFERENCES

BEGNI, G., 1982 Selection of the optimum spectral

bands for the SPOT satellite. Photogrammetric Engineering and Remote Sensing, Vol.48, No.10, pp1613-1620.

CNES, SPOT-IMAGE, 1986 Guide des utilisateurs de données SPOT. SI/MR/86.005F.

PROY, C., 1986 Integration du relief au traitement d'images de télédétection. D.I. Thesis Institut National Polytechnique, Toulouse, France.

WOODHAM, R.J., GRAY, M.H., 1987 An analytic method for radiometric correction of satellite multispectral scanner data, IEEE Transactions on geoscience and remote sensing. VOL.GE-25, No.3.

YANG, C.J., VIDAL, A., 1989 Removal of the scan angle effect on NOAA-AVHRR data in visible and near-IR parts of the electromagnetic spectrum. IGARSS'89.



FIGURE 1

Digital Elevation Model image. Dark areas represent low elevation, light areas represent high elevation.



FIGURE 2

Slope image derived from Digital Elevation Model.



FIGURE 3

$\cos I$ image generated using equation (4).



FIGURE 4

A portion of SPOT-1 HRV image (band 1) of Picaussel Forest.



FIGURE 5

Synthetic reflectance factor image generated using equation (3).

COMPUTER ASSISTED LAND COVER MAPPING WITH SPOT IN INDONESIA

Jean-Philippe GASTELLU-ETCHEGORRY
PUSPICS, Gadjah Mada University, Yogyakarta, Indonesia
SCOT, Earth Observation Consultancy Service, Toulouse, France.

Daniele DUCROS-GAMBART
C.E.S.R., Paul SABATIER University
9, avenue du Colonel Roche - BP 4346
31029 Toulouse Cedex France

The capability of SPOT combined with specifically designed classifiers was investigated for computer assisted land cover/use mapping in Indonesia. Two particularly disturbing constraints were analysed. (1) The atmospheric upwelling radiances may have very large values (up to 80% of the measured radiance for band XS1) and may be characterized by an important heterogeneity (up to 40% for all bands within a 60 x 60 km SPOT image). (2) Moreover, the small size, complexity and dynamic nature of Indonesian agro-forest systems confuse multispectral analysis.

In that case, conventional spectral classifiers are inadequate; texture and context information must be introduced in order to derive both adequate and accurate information. Consequently, two different methodologies were investigated and developed.

The first method consists in classifying a large number of spectral classes/sub-classes; this number must be sufficiently large for obtaining small within-class variances. Thanks to visual information these classes are later on grouped according to local context and texture. With the objective of making computer land cover mapping more objective and efficient it was also developed a layered classifier combined with an automatic exploitation of local textural and contextual information. Both methods are based on micro-computer technology in order to obtain systems that can be readily operational in Indonesia.

These classifiers were tested on two study areas in Java. Preliminary results suggest that in Indonesia SPOT has the potential of a major data source for deriving land cover mapping to 1:50,000/100,000 scales.

With the first method, six main land cover/use classes (lake, rice field, dry crop, settlement/road, mixed garden and forest/plantation) were obtained. Eighteen significative classes were discriminated with the second method.

The accuracy of the SPOT land cover / use maps was tested in the field through random sampling of all land cover classes. Generally speaking, it was found 65%-100% probability that these are rightly classified (according to the classes).

OPTIMISATION DE LA RESTITUTION AUTOMATIQUE DU RELIEF À PARTIR D'IMAGES SPOT

A. JATON, R. SIMARD, A. LECLERC ET S.R. HAJA

Digim inc.
1100, boul. René-Lévesque Ouest
Montréal (Québec)
H3B 4P3

En collaboration avec le
Centre d'applications et de recherches en télédétection (CARTEL)
Université de Sherbrooke
Sherbrooke (Québec)
J1K 2R1

RÉSUMÉ

L'objet de cette recherche est la compréhension des facteurs qui influencent la qualité des modèles numériques d'élévation produits par corrélation d'images SPOT stéréoscopiques.

De nombreux tests ont permis l'évaluation des liens entre la précision obtenue à l'appariement des images et les facteurs suivants : la radiométrie des images, les distorsions dues au relief, les grandeurs des fenêtres d'appariement et les critères d'acceptation des mesures de parallaxe.

Les résultats montrent l'importance de la variance des images comme critère de sélection d'un bon appariement et permettent de juger du rôle du coefficient de corrélation. Finalement, une évaluation des effets géométriques sur les erreurs de corrélation est effectuée, précisant ainsi le rapport entre la grandeur des fenêtres d'appariement et les distorsions dues au relief.

Dans des études antérieures, Guindon (1986) s'est penché sur le problème du critère de sélection d'un bon appariement avec des images radars. Il en conclut que la robustesse du coefficient de corrélation est très sérieusement mise en doute comme bon critère d'acceptation et que la variance des images a un avenir prometteur à ce titre. En ce qui a trait aux grandeurs des fenêtres d'appariement, Rosenholm (1987) aborde le fait qu'il existe, pour tous types de terrain, une grandeur de fenêtre optimale. Son étude basée sur l'appariement de photos aériennes par moindres carrés, indique que les fenêtres à utiliser devraient avoir une dimension variant de 20x20 à 30x30 pixels.

Notre étude vise à poursuivre les recherches précédentes et à expliquer de façon plus détaillée les facteurs en cause en nous basant sur de vraies données SPOT. D'ailleurs, la compréhension des paramètres influents est primordiale si l'on veut réaliser une cartographie topographique complètement automatisée et précise.

1. INTRODUCTION

Depuis l'avènement des images satellites et principalement depuis le lancement du satellite SPOT, la restitution du relief s'est nettement orientée vers les algorithmes automatiques d'appariement d'images numériques. Bien des algorithmes ont été élaborés afin de générer des modèles numériques d'élévation (MNE) (Simard, 1981; Rochon et al., 1984; Cooper et al., 1985; Ehlers et al., 1987). Les résultats sont très encourageants et la précision des cartes topographiques résultantes atteint les normes cartographiques (environ 3 à 5 mètres de précision e. altimétrie). Par contre, aucune méthode ne peut se vanter de produire des MNE de façon complètement automatisée sans l'intervention d'un opérateur.

L'élimination de ces erreurs mais surtout la compréhension des facteurs influençant l'appariement d'images ont été des sujets de recherche pour de nombreux scientifiques. Les problèmes causés par la radiométrie des images, les distorsions géométriques et le choix des grandeurs de fenêtre d'appariement sont de taille, et pire encore, sont intimement liés. Une autre source importante de questions s'ajoute dans notre cas : le rôle du coefficient de corrélation comme mesure de fiabilité des valeurs de parallaxe obtenues.

2. MÉTHODOLOGIE

2.1 L'algorithme de corrélation normalisée

L'algorithme de corrélation normalisée est un module du système CARTOSPOT (Leclerc, 1988; Rochon et al., 1984). Il nécessite deux images SPOT (P ou XS) rééchantillonnées dans le système intermédiaire moyen (SIM). Le SIM assure la correction des images de toutes les déformations géométriques autres que celles dues au relief. Il permet également de diminuer très fortement la parallaxe en y. Le rééchantillonnage des images se fait à l'aide des paramètres de la modélisation géométrique. Le modèle utilisé est celui de Toutin et Guichard (Toutin, 1985) et nécessite des points d'appui, les paramètres orbitaux et d'attitude et les données d'informations générales sur les images.

Le logiciel d'appariement exige les paramètres suivants :

- la grandeur minimale de la fenêtre de référence;
- la grandeur minimale de la fenêtre de recherche;
- le seuil de variance radiométrique dans la fenêtre de référence;
- le seuil de corrélation;
- les positions respectives d'un point dans les deux images pour initialiser le processus.

Pour chaque pixel de l'image de référence, un premier test doit être réussi : le test de variance. Si la fenêtre entourant ce pixel ne contient pas une variance supérieure au seuil indiqué, la grandeur de la fenêtre est augmentée jusqu'à ce que l'un des deux phénomènes suivants se produise : le rejet du point si la fenêtre a été agrandie jusqu'à son maximum de 16x16 sans que sa variance atteigne le seuil requis, ou le test de variance est concluant et le processus se poursuit. Dans ce dernier cas, la mesure du coefficient de corrélation normalisé est effectuée pour chacune des positions de la fenêtre conjuguée dans la zone de recherche. Si cette mesure dépasse le seuil spécifié, la position fractionnaire exacte du point homologue de l'image conjuguée est interpolée en localisant le sommet de la fonction de corrélation à l'aide des trois valeurs maximales. Puis, la soustraction des positions de l'image de référence de celles de l'image conjuguée fournit la parallaxe.

Après filtrage de la parallaxe afin d'éliminer les erreurs grossières, les valeurs d'élévation sont déduites à l'aide des paramètres de la modélisation géométrique. Précisons que l'obtention de l'altitude peut également se faire de façon hiérarchique, en utilisant des images de moindre résolution spatiale. L'ajout d'information, en augmentant la résolution dans un processus itératif, permet d'obtenir des modèles de terrain plus robustes et précis.

2.2 Les données images

Les images utilisées pour notre étude sont des images stéréoscopiques SPOT panchromatiques de la Malaysia dont les pixels sont réduits à 50 mètres de résolution (par moyenne sur des fenêtres 5x5). Ces images ont été acquises les 13 et 17 février 1987 et présentent un rapport base-hauteur de 0,87. Afin de comparer la validité des résultats d'appariement, un fichier de parallaxe a été corrigé visuellement (à partir des ortho-images en superposition). Ce fichier a servi de référence afin de déceler tout particulièrement les erreurs grossières. Les images complètement corrigées de l'effet du relief (ortho-images) interviennent également dans les traitements.

2.3 Méthodologie utilisée

Le travail est divisé en trois étapes ayant pour but de comprendre respectivement : le rôle du coefficient de corrélation, l'importance de la radiométrie et les effets des distorsions géométriques. Ces étapes comprennent chacune le découpage de régions tests, les appariements d'images, la recherche et la compilation des erreurs, ainsi que l'analyse des résultats.

3. RÔLE DU COEFFICIENT DE CORRÉLATION

3.1 Tests effectués

Des appariements avec une grandeur de fenêtre et un seuil de variance fixes mais un seuil de corrélation variable ont été effectués sur deux sous-images de la Malaysia de 100 lignes par 100 pixels. L'une de ces sous-images (zone A) contient une zone montagneuse couverte de forêt tropicale à variance faible et l'autre (zone B) présente un terrain plat que recouvrent un village et des cultures (variance élevée). Après les corrélations et comparaisons avec le fichier de parallaxes de référence, les pourcentages d'erreurs grossières (définies comme

étant, dans ce cas, supérieures à 2,5 fois la dimension du pixel, c'est-à-dire 125 mètres) ainsi que le pourcentage de pixels pour lesquels la corrélation a été infructueuse sont calculés.

3.2 Résultats

Le tableau ci-dessous présente, pour chacune des zones-tests, une partie des résultats obtenus :

Tableau 1 : Les erreurs par rapport au coefficient de corrélation.

| ZONE A | | | | | | ZONE B | | | | | |
|--------|-----|----|-------|--------|-------|--------|-------|--------|-------|--------|--|
| Sc | Fc | Sv | Fr=10 | | Fr=16 | | Fr=10 | | Fr=16 | | |
| | | | % err | % n.c. | % err | % n.c. | % err | % n.c. | % err | % n.c. | |
| 0,2 | 3x9 | 4 | 0,64 | 7,14 | 0,01 | 6,58 | 0,03 | 0,04 | 0,21 | 0,80 | |
| 0,4 | 3x9 | 4 | 0,03 | 54,75 | 0,00 | 78,33 | 0,00 | 0,68 | 0,00 | 2,53 | |
| 0,6 | 3x9 | 4 | 0,00 | 97,34 | 0,00 | 100,00 | 0,00 | 4,24 | 0,00 | 4,58 | |
| 0,8 | 3x9 | 4 | 0,00 | 100,00 | 0,00 | 100,00 | 0,00 | 61,64 | 0,00 | 55,22 | |

où : Sc = seuil de corrélation;
Fc = grandeur de la zone de recherche dans la seconde image;
Sv = seuil de variance radiométrique;
Fr = grandeur de la fenêtre de référence;
% err = pourcentage d'erreurs supérieures à 2,5 pixels de parallaxe;
% n.c. = pourcentage de pixels n'ayant pas été corrélés.

3.3 Analyse : Le coefficient de corrélation n'est pas un bon critère d'appariement.

Les résultats montrent de façon frappante que la valeur du coefficient de corrélation n'est pas significative. En effet, dans les deux zones, si le seuil était fixé à 0,6 et la grandeur de fenêtre de référence à 16, dans le premier cas aucun pixel n'aurait trouvé son homologue, alors que dans le second, 95% l'auraient trouvé sans erreur! De plus, alors que statistiquement une corrélation est bonne lorsque son coefficient atteint 0,7, la zone A, malgré un seuil aussi faible que 0,2, présente des résultats fort acceptables. En fait, il faut considérer le coefficient de façon relative et non absolue pour l'appariement, puisque seule la courbe de corrélation a une grande importance.

4. LA RADIOMÉTRIE DES IMAGES

4.1 Tests effectués

Dans le but de comprendre l'effet et l'importance de la variance des images, il a fallu en dissocier les effets provenant de la géométrie. Ainsi, afin d'éliminer les erreurs provenant des effets géométriques, les appariements ont été effectués sur des ortho-images. Trois sous-images de variance homogène faible, moyenne et forte ont été extraites des ortho-images et appariées avec des grandeurs de fenêtre variables. A nouveau, la proportion des erreurs et des pixels non corrélés est évaluée. Notons que puisqu'il s'agit d'ortho-images et que, par conséquent, les erreurs sont moins importantes, seules les erreurs supérieures à 0,5 et 1 pixel de parallaxe sont comptabilisées.

4.2 Résultats

Le tableau 2 ci-dessous présente, pour trois zones (zone A, B et C) de variance différente, les erreurs de corrélation pour différentes grandeurs de fenêtre et de tolérance pour la détection des erreurs. La figure 1 en résume le contenu.

Tableau 2 : Erreurs supérieures à 0,5 (a) et 1 (b) pixel de parallaxe (%) par rapport aux variations de la grandeur de la fenêtre de corrélation et de la variance. Les zones A, B et C présentent un écart-type radiométrique moyen de 5,7, 15,2 et 38,5 en niveaux de gris respectivement.

| | Fr | 6 | 8 | 10 | 12 | 14 | 16 |
|-----------|-------|-------|-------|-------|-------|-------|----|
| a) zone A | 42,78 | 39,76 | 39,19 | 38,74 | 38,35 | 38,10 | |
| zone B | 28,36 | 25,05 | 23,16 | 21,56 | 20,38 | 19,17 | |
| zone C | 21,13 | 14,67 | 10,67 | 7,06 | 4,60 | 2,74 | |
| b) zone A | 6,62 | 1,47 | 0,56 | 0,02 | 0,00 | 0,00 | |
| zone B | 0,79 | 0,21 | 0,00 | 0,00 | 0,00 | 0,00 | |
| zone C | 0,29 | 0,13 | 0,00 | 0,00 | 0,00 | 0,00 | |

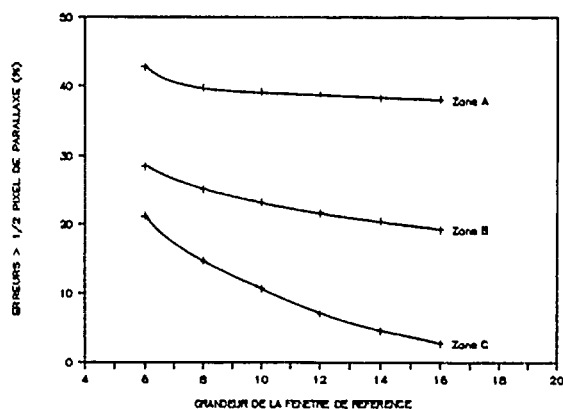


Figure 1 : Les erreurs de corrélation en fonction de la grandeur de la fenêtre de référence pour trois types de terrain. Les zones A, B et C présentent un écart-type radiométrique moyen de respectivement 5,7, 15,2 et 38,5 en niveaux de gris.

4.3 Analyse : La variance joue un rôle primordial lors de la corrélation.

Les résultats s'améliorent toujours avec l'augmentation de la grandeur de la fenêtre de référence et donc, de la variance. De plus, plus la variance d'une zone est forte, meilleurs sont les résultats, et ceci même avec une dimension de fenêtre restreinte. Notons également l'importance de l'amélioration de la précision sur la zone A (6%) lorsque la fenêtre passe de 6x6 à 16x16 (tableau 2, cas b). C'est énorme si l'on considère que ce chiffre (6%) représente 2 160 000 pixels sur une image SPOT! Ainsi comme le pressentait Guindon, la variance peut jouer un rôle primordial comme critère de sélection d'un bon appariement et supplanter le coefficient de corrélation.

5. LES DISTORSIONS GÉOMÉTRIQUES

5.1 Tests effectués

Afin de ne considérer à cette étape que l'effet dû au relief, des couples stéréoscopiques sont simulés. À partir d'une ortho-image et d'un MNE réels, une visée oblique simulante la seconde image avec l'angle désiré a été créée (Simard, 1981). Trois ortho-images et deux MNE ont donc servi à générer six couples stéréoscopiques de rapport base-hauteur égal à 0,466. Les deux MNE se distinguent par la force de leur pente. Le MNE A présente des variations d'altitude de 150 m en moyenne et le MNE B, de 95 m. Les appariements sont donc réalisés sur des images ne présentant que des distorsions dues au relief et ayant la même radiométrie d'origine.

5.2 Résultats

Vu la similitude des résultats pour les trois couples d'images, seuls les résultats d'une sous-région sont énoncés. Ces résultats se retrouvent dans le tableau 3 et à la figure 2.

Tableau 3 : Les erreurs induites par les distorsions géométriques.
a) Distorsions géométriques provenant du MNE A
b) Distorsions géométriques provenant du MNE B

| | Fr | 6 | 8 | 10 | 12 | 14 | 16 |
|--------------|-------|-------|-------|-------|-------|-------|----|
| a) Err 1 (%) | 1,37 | 2,29 | 4,16 | 5,74 | 8,03 | 10,78 | |
| Err 2 (%) | 0,10 | 0,13 | 0,15 | 0,07 | 0,14 | 0,34 | |
| Err moy (m) | 15,57 | 16,40 | 18,26 | 20,17 | 22,07 | 24,15 | |
| b) Err 1 (%) | 4,23 | 5,69 | 8,58 | 11,69 | 14,95 | 19,23 | |
| Err 2 (%) | 0,13 | 0,07 | 0,16 | 0,29 | 0,43 | 0,72 | |
| Err moy (m) | 18,16 | 20,16 | 22,83 | 25,32 | 27,94 | 30,68 | |

où: Err 1 = Pourcentage d'erreurs supérieures à 0,5 pixel de parallaxe (53 m en altimétrie).
Err 2 = Pourcentage d'erreurs supérieures à 1 pixel de parallaxe (107 m en altimétrie);
Err moy = Moyenne des valeurs absolues des erreurs d'élévation.

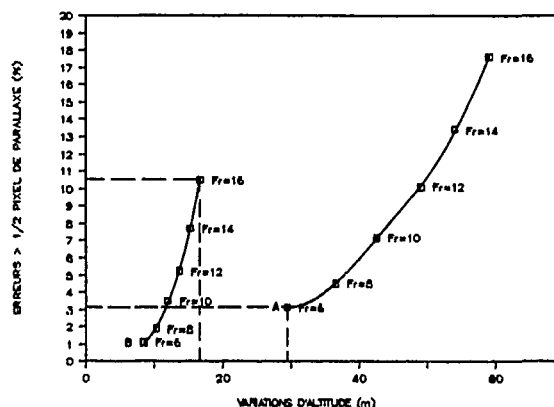


Figure 2 : Les erreurs en fonction des variations d'altitude (écart-type) du terrain à l'intérieur de la fenêtre Fr pour le MNE A (A) et le MNE B (B).

5.3 Analyse : La corrélation avec de petites fenêtres est beaucoup moins sensible que prévu aux effets géométriques.

En premier lieu, remarquons que l'agrandissement de la fenêtre de référence (et donc des distorsions géométriques à l'intérieur de celle-ci) provoque dans tous les cas une augmentation des erreurs. Ensuite, il est très intéressant de noter que, par exemple, corréler deux images présentant des distorsions géométriques moyennes (MNE B) avec une grande fenêtre de 16x16 engendre beaucoup plus d'erreurs que la corrélation de deux images ayant subi des distorsions fortes (MNE A) avec une petite fenêtre de 6x6 (figure 2). Ce résultat indique nécessairement que la corrélation avec de petites fenêtres s'accommode très bien de fortes distorsions géométriques. On pourrait avancer que l'auto-corrélation des images et la structure du calcul des coefficients de corrélation normalisés sont en relation étroite avec ce phénomène. Afin de vérifier cette hypothèse, un modèle sera élaboré sous peu et testé avec ces mêmes données.

6. AMÉLIORATIONS FUTURES DU LOGICIEL

6.1 Adaptation locale du logiciel à la variance

Puisque le coefficient de corrélation ne peut servir qu'à l'élimination de très fortes erreurs et qu'il faut, vu les résultats le garder très bas, la variance peut alors être d'un bon secours comme filtre des erreurs. Le seuil sur la variance déjà installé dans le logiciel servirait à éliminer les pixels ne présentant pas, pour tout type de fenêtre, une valeur minimale de variance. Une fois le test sur la variance minimale réussi, le choix de la fenêtre optimale pourrait se faire à partir d'une table associant la grandeur de la fenêtre à la variance locale. La définition de cette table sera complétée aussitôt que l'analyse des présents tests sera terminée. Notons à ce sujet que les erreurs provoquées par une faible variance sont beaucoup plus importantes que celles provenant des distorsions géométriques et que, par conséquent, il faut s'en occuper de façon prioritaire.

6.2 Adaptation du logiciel au relief

Le logiciel tel qu'il est utilisé maintenant permet la production de MNE de façon itérative par un processus appelé stéréoréduction. Ce processus, qu'il convient de préciser ici, donne la possibilité de produire un MNE en plusieurs étapes et est utilisé présentement dans les régions de relief accidenté. La première étape fournit un MNE grossier, alors que les suivantes apportent la précision manquante. Ce processus est intéressant dans notre cas puisque, à résolution spatiale réduite, le rapport signal-bruit est supérieur et les distorsions géométriques beaucoup moins prononcées. Les précisions apportées par cette étude sur les effets des distorsions géométriques permettront certainement d'utiliser plus efficacement la stéréoréduction. Comme il a été noté, la corrélation avec de petites fenêtres est assez robuste face aux forts reliefs. Il n'est donc pas rentable de trop utiliser la stéréoréduction lorsque la qualité radiométrique des images le permet.

L'achèvement des tests et de l'analyse des résultats permettra de construire un modèle fournissant, pour une image donnée (c'est-à-dire pour une certaine variance moyenne), un terrain donné et bien sûr un temps maximum de traitement, le nombre d'étapes de stéréoréduction à utiliser et la précision attendue.

7. CONCLUSION

Les résultats de cette étude apportent un peu de lumière sur les paramètres importants lors de la restitution automatique du relief obtenue à partir de corrélations numériques d'images. Ils mettent d'abord en relief le peu de poids que possède le coefficient de corrélation comme mesure de la fiabilité des appariements. L'emphasis doit plutôt être mise sur l'allure de la courbe de corrélation, puisque, dans certaines zones, une faible valeur du coefficient de corrélation peut s'avérer acceptable.

Du côté de la radiométrie, l'importance de la variance des images est cruciale. Les erreurs sont directement liées aux faibles variances. Ceci signifie que plus les dimensions des fenêtres sont grandes, meilleurs sont les résultats de corrélation.

Du point de vue géométrique, par contre, les erreurs sont minimisées lors de l'utilisation d'une petite fenêtre et s'accroissent lorsque cette dernière est agrandie. Ainsi existe-t-il une grandeur de fenêtre

optimale qui se situe entre 8x8 et 16x16 dépendamment du type de terrain. Plus intéressant encore est le fait qu'un MNE généré par une corrélation avec une petite fenêtre sur un terrain pentu est de meilleure qualité qu'avec une corrélation avec une grande fenêtre sur un terrain moins pentu (bien sûr, il y a une limite calculable à l'étendue de ce phénomène).

Ces résultats concrets et ces conclusions simples vont par conséquent être fort utiles pour l'adaptation du logiciel existant à la fois à la radiométrie et à la géométrie des images. Le logiciel utilisant déjà un seuil sur la variance, une grandeur de fenêtre variable et la possibilité de stéréoréduction, il sera possible d'accroître, avec un effort de programmation limité, l'automatisation du processus et la précision des MNE produits.

Finalement, ces résultats confirment la pertinence des choix initiaux de Digim concernant le développement du système CARTOSPOT. Parmi l'ensemble des méthodes d'appariement d'images, la corrélation normalisée, avec adaptation de la grandeur des fenêtres et combinée à la méthode hiérarchique de stéréoréduction, est bien adaptée aux conditions rencontrées en cartographie topographique d'environnements naturels caractérisés par des déformations géométriques, des contrastes et des textures d'images variables.

RÉFÉRENCES

1. Cooper, P.R., D.E. Friedmann et S.A. Wood, "The Automatic Generation of Digital Terrain Models from Satellite Images by Stereo", Comptes rendus du 36th Congress of the International Astronautical Federation, du 7 au 12 octobre, Stockholm, Suède, 1985.
2. Ehlers, M. et R. Welch, "Stereocorrelation of LANDSAT TM Images", Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 9, p. 1231-1237, 1987.
3. Guindon, B. et H. Maruyama, "Automated Matching of Real and Simulated SAR Imagery as a Tool for Ground Control Point Acquisition", Journal canadien de télédétection, décembre 1986.
4. Leclerc, A., "Le système CARTOSPOT", Actes du symposium international sur les applications topographiques des données SPOT, 13 et 14 octobre, Sherbrooke, Québec, 1988.
5. Rochon, G., A. Leclerc, S.R. Haja et R. Simard, "La création de modèles numériques de terrain à l'aide d'images des Satellites SPOT", Comptes rendus du 9^e symposium canadien sur la télédétection, Saint-Jean, Terre-Neuve, 1984.
6. Rosenholm, D., "Empirical Investigation of Optimal Window Size using the Least Squares Image Matching Method", Photogrammetria, Vol. 42, pp113-125, 1987.
7. Simard, R., "Résultats de simulations d'images stéréoscopiques HRV SPOT sur le site de Gun Lake, C.-B.", Comptes rendus du 7^e symposium canadien sur la télédétection, du 8 au 11 novembre, Winnipeg, Manitoba, 1981.
8. Toutin, Th., "Étude mathématique pour la rectification d'images SPOT", Comptes rendus du 26^e congrès de la Fédération internationale des géomètres, du 1^{er} au 11 juin, Toronto, Ontario, 1986.

APPROCHE MULTITEMPORELLE ET TEXTURALE DES DONNEES SPOT MULTISPECTRALES POUR LE SUIVI AGRICOLE EN AFRIQUE CENTRALE.

P. Jacques, M. Massart, J. Wilmet.

Laboratoire de Télédétection
Place Louis Pasteur, 3 - 1348 Louvain-La-Neuve, Belgique
Tel : 10/472870 Telex : 59037 uclb Fax : 32 10 472999

RESUME

La méthode présentée permet de faciliter la discrimination des champs traditionnels en ajoutant à l'information spectrale des images satellitaires SPOT, l'information spatiale représentée dans les matrices de co-occurrence. Une méthodologie générale d'utilisation des différentes plate-formes satellitaires dans un but d'obtention de statistiques agricoles au Shaba (Zaire) est présentée. La faisabilité et les coûts sont analysés.

ABSTRACT

In order to ease the discrimination of agricultural fields in humid tropical areas, the spatial information present in cooccurrence matrices is added to the spectral information of a satellite image. A general methodology for combining satellite platforms to obtain agricultural statistics in Shaba (Zaire) is presented. The feasibility and the costs are analyzed.

MOTS-CLES

Suivi agricole, Afrique, Zaire, Statistiques agricoles, Satellite SPOT.

1. INTRODUCTION

S'inscrivant dans le cadre du projet TELSAT 09 du Service de la Programmation de la Politique Belge, ayant pour objectif "l'Amélioration des Statistiques Agricoles par Télédétection en République du Zaire", cette étude a pour objectif :

- L'estimation des potentialités des satellites haute résolution dans la discrimination des principales espèces cultivées en milieu tropical humide. Cette première approche devrait permettre l'élaboration d'une méthodologie pour l'évaluation des superficies mises en culture par espèces.
- La mise au point d'un système opérationnel de collecte des Statistiques Agricoles en milieu tropical par une approche intégrée des différents capteurs satellitaires.

L'aire d'étude, retenue pour ses caractéristiques agricoles, se situe au Shaba, au nord-ouest de Lubumbashi, dans la vallée de la Moyenne Lufira. L'aire ainsi définie, centrée sur la collectivité des Bayeke (cpt Bunkeya), a une superficie de plus de 100 km².

Son aspect général témoigne des caractéristiques agricoles traditionnelles du milieu shabien : parcelles de faibles dimensions (inférieures à 50 ares) concentrées en des sites de culture bien particuliers; ainsi on observe peu de parcelles de culture isolées au sein de la végétation naturelle.

La région est caractérisée par un climat tropical à saison sèche unique. La saison agricole débute en octobre avec les premières pluies pour se terminer en avril avec la fin de la saison humide.

Les principales espèces cultivées sont le maïs et le manioc. On observe de plus un certain nombre de parcelles de cultures maraichères de contre-saison localisées dans les endroits humides, sur alluvions en bordure de rivières.

En ce qui concerne la végétation naturelle, on constate une large interpénétration des différents types de formations végétales : forêt dense sèche (muhulu), forêt galerie, forêt claire (miombo), savane arborée, arbustive ou herbeuse.

2. METHODES

2.1 DONNEES DISPONIBLES

2.1.1 Données satellitaires

Trois images multispectrales -scene SPOT 123/371- sont à disposition, elles sont datées respectivement du 30/07/1987, du 26/04/88 et du 11/05/88. Nous pouvons encore bénéficier d'une image panchromatique enregistrée le 23/12/87.

2.1.2 Données terrain

Un contrôle de terrain a été réalisé en mai 1988, soit quelques jours après l'enregistrement satellitaire. Il s'est appuyé sur un survol aérien en monomoteur effectué préalablement à l'analyse au sol.

Celle-ci a permis la constitution d'une ventée terrain par investigation au niveau du parcellaire.

Un report des parcelles investiguées a été effectué sur l'image panchromatique de décembre. L'observation de terrain a conduit à la définition des aires d'entraînement nécessaires à la classification supervisée ainsi qu'à la préparation d'un sondage supplémentaire destiné à vérifier la précision des classifications par matrices de contingence.

2.2 PRELIMINAIRES

Etant donné le caractère opérationnel recherché, c'est-à-dire l'obtention de statistiques agricoles pour l'ensemble du Shaba, un schéma de travail restrictif a été adopté :

a. L'étude se concentrera sur l'analyse et l'extraction optimale de l'information contenue au niveau d'une seule image satellite par saison agricole.

Ce choix est justifié par :

- Une analyse de la probabilité d'obtention d'une couverture satellitaire totale du Shaba au cours d'une saison agricole. Si cette probabilité est d'environ 80% pour une couverture totale, elle s'abaisse à $0.8 \times 0.8 = 0.64$ soit 64% lorsque la demande est de deux couvertures.
 - Un coût d'acquisition et de traitement des données proportionnel au nombre de couvertures satellitaires par saison agricole souhaité.
- b. Les traitements longs et complexes seront évités, privilégiant une approche des données simple et directe.

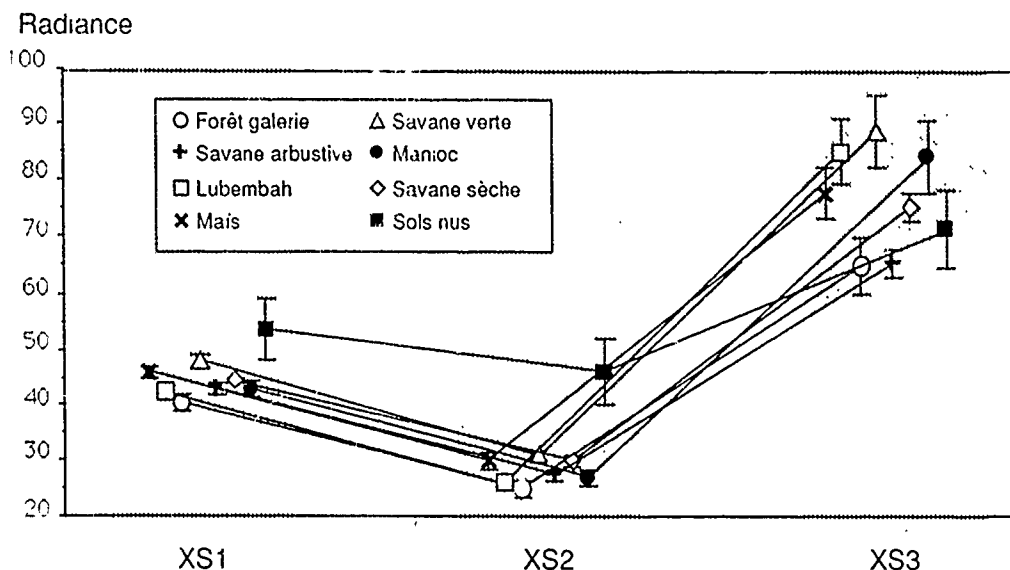
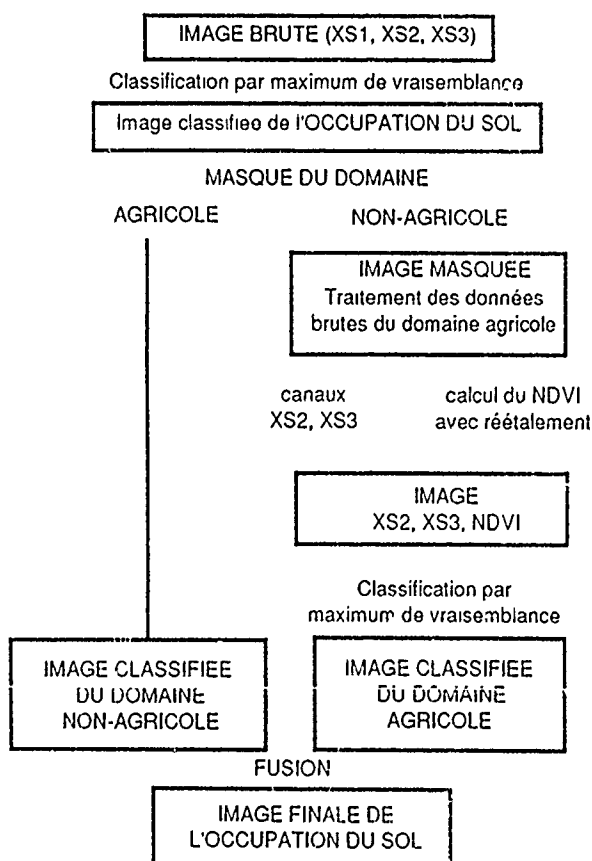


Figure 1 Valeurs moyennes et écarts types des signatures spectrales des taxons identifiés dans l'aire d'étude.

Figure 2 Schéma de traitement de l'image utilisant le néo-canal NDVI réévalué sur les données du domaine agricole



2.3 ANALYSE DES DONNEES

Compte tenu des contraintes énoncées précédemment, la période optimale d'acquisition des données se situe en mars et en avril. A ce moment, toutes les cultures sont en place, que ce soient les cultures de saison des pluies (manioc et maïs proches de la récolte) ou les parcelles maraichères (en préparation dès avril). Cette période correspondant à la fin de la saison des pluies, se caractérise de plus, par une diminution notable de la couverture nuageuse, et précède une hausse de la nébulosité atmosphérique due en saison sèche aux feux de brousse et poussières atmosphériques.

Pour l'année agricole 1987-1988, nous disposons de trois enregistrements sur le site d'étude retenu.

- L'image panchromatique de décembre présente une forte couverture nuageuse, principalement sur le site test de Bunkeya, ce qui limite son utilisation pratique.
- L'image multispectrale de mai présente, elle, une forte nébulosité en des zones localisées due aux feux de brousse, ce qui introduit un biais difficilement conciliable avec des traitements à plus petite échelle.
- L'image d'avril sera donc seule retenue pour une première analyse.

2.4 TRAITEMENTS

Les taxons les plus importants sur l'image d'avril 1988 sont difficilement différenciables avec les méthodes habituelles de classification (figure 1). En particulier, la méthode du maximum de vraisemblance est insuffisante pour discriminer les taxons liés à l'agriculture. Une première étude utilisant la bipolarisation (Delouin et al., 1987) a pallié ce problème mais reste difficile à mettre en oeuvre. Une procédure exploitant l'information spécifique apportée par l'indice de la Végétation de la Différence Normalisée (NDVI) et optimisant cet apport au niveau des superficies agricoles a été tentée. Elle procède par masquage et réévaluation du néo-canal pour finalement conduire à une classification plus précise des taxons qui constituent l'objet de l'étude (figure 2). Cette méthode avait déjà été appliquée avec succès dans un environnement de type soudano-sahélien au Burkina-Faso (Lambin, 1988). Les résultats obtenus au Sud-Shaba sont assez satisfaisants mais malheureusement restent subordonnés à une définition précise d'aires d'entraînement.

La méthode que nous présentons par la suite est plus simple, plus rapide et donc plus facile à exécuter dans un processus opérationnel.

Un premier traitement consiste en une classification par maximum de vraisemblance sur l'image à six canaux résultat de la superposition de l'image brute d'avril 1988 et de l'image de juillet 1987 corrigée géométriquement par rapport à la précédente.

Trois classes sont extraites :

- les sols nus, villages et pistes communs aux deux dates,
- la rivière et les zones humides,
- le lubembah (zones marécageuses)

L'image de base est masquée pour ces trois classes, ce qui permet dans un premier temps d'enlever la classe lubembah qui se confond avec la classe manioc.

Une étude spatiale introduit un facteur supplémentaire pour la différenciation entre les zones de culture et la végétation environnante. En effet, l'analyse de terrain a montré le fort contraste existant entre les zones de culture fortement entropiques, concentration et juxtaposition de champs de petite taille et de réflectance différente (maïs, manioc et sols nus), et la végétation environnante plus homogène.

A cette fin, nous introduisons les matrices de co-occurrence et les paramètres de texture.

Soit une image $I(x, y)$, $1 \leq x, y \leq n$

avec $0 \leq I(x, y) \leq k$ k étant le niveau de gris maximum.

La matrice de co-occurrence C_{Δ} pour cette image au déplacement :

$\Delta = (\Delta x, \Delta y)$ est définie par :

$$C_{\Delta}(i_1, i_2) = \frac{\#\{(x, y), (x+\Delta x, y+\Delta y) : I(x, y) = i_1, I(x+\Delta x, y+\Delta y) = i_2\}}{\#\{(x, y), (x+\Delta x, y+\Delta y) : 1 \leq x, y, x+\Delta x, y+\Delta y \leq n\}}$$

où $\#S$ est le nombre d'éléments dans l'ensemble S .

La matrice de co-occurrence représente donc une estimation de la probabilité que la paire de niveaux de gris (i_1, i_2) soit trouvée en des pixels distants de Δ .

Haralick (Haralick et al, 1973) introduit plusieurs paramètres caractérisant la texture et notamment :

$$\text{Le contraste } C = \sum_{i=1}^k \sum_{j=1}^k (i-j)^2 C_{\Delta}(i, j)$$

$$\text{L'entropie } E = - \sum_{i=1}^k \sum_{j=1}^k C_{\Delta}(i, j) \log(C_{\Delta}(i, j))$$

$$\text{L'homogénéité } H = \sum_{i=1}^k \sum_{j=1}^k \{C_{\Delta}(i, j)\}^2$$

Les trois paramètres contraste, entropie, homogénéité sont calculés, dans une matrice de convolution 3×3 , sur chacun des canaux de l'image d'avril 88. Le résultat est une image contenant neuf nouveaux néo-canaux spectraux reflétant les relations spatiales entre les pixels.

Plusieurs constatations vont nous permettre de réduire la quantité de données produites.

- le paramètre entropie fortement corrélé au paramètre homogénéité peut être éliminé,
- les paramètres calculés sur le canal XS3 reflètent le mieux l'hétérogénéité spatiale des zones de culture

Les deux néo-canaux "homogénéité" et "contraste" calculés sur le canal XS3 sont ajoutés aux trois canaux de l'image d'avril. Les étapes de la procédure sont les suivantes.

- 1) Une première classification par hypercubes sur cette image à cinq canaux masquée afin d'extraire la classe manioc. La classification par maximum de vraisemblance n'est plus applicable, la distribution des classes n'étant plus gaussienne.
- 2) Masquage de l'image par la classification
- 3) Classification par hypercubes sur l'image à cinq canaux masquée pour obtenir la classe maïs traditionnel
- 4) Classification par hypercubes sur l'image de base à trois canaux masquée pour obtenir le maïs industriel.

3 RESULTATS

| | Nombre de pixels | Surfaces (hectares) |
|-------------------|------------------|---------------------|
| Manioc | 4729 | 189,16 |
| Maïs traditionnel | 13407 | 536,28 |
| Maïs industriel | 1924 | 76,96 |

Le développement méthodologique mis au point a permis la discrimination des parcelles des principales espèces cultivées. Dans l'étude des données SPOT multispectrales, on conclut à une fiabilité de près de 80% dans l'estimation des superficies cultivées par espèces (manioc, maïs). Les résultats ont été évalués par photographies aériennes basse altitude.

4 ANALYSE ECONOMIQUE

L'extrapolation de la méthodologie au niveau régional shabien ne passe plus simplement par une analyse de faisabilité mais demande une étude précise des coûts relatifs à l'obtention et au traitement des données.

Dans un souci d'opérationnalité et de réduction des coûts, un schéma de travail s'appuyant sur le sondage stratifié - stratification sous-régionale préalable - semble nécessaire.

La variance de l'estimation sera d'autant plus réduite que l'échantillonnage se trouvera bien adapté aux différents milieux agricoles rencontrés. C'est à cet effet qu'une stratification préalable est requise. La stratification effectuée sur données LANDSAT MSS est réalisée principalement sur base de critères d'occupation du sol, prépondérance et densité de certaines cultures, ainsi que suivant les conditions agro-pédo-climatologiques. L'échantillonnage se devra donc d'être adapté dans sa densité à l'affectation du sol. Etant donné la relative stabilité du milieu, un traitement d'enregistrements datant de quelques années (six à sept ans) reste acceptable dans la définition du domaine agricole et des grandes formations naturelles. L'analyse numérique des données CCT MSS pourra encore conduire à ne conserver que le domaine agricole qui sera défini plus profondément dans des études ultérieures. Les zones n'ayant aucune culture d'intérêt seront regroupées au sein d'une même strate qui ne fera l'objet d'aucune enquête et ne nécessitera donc pas l'achat de données haute résolution pour une investigation plus fine.

C'est au niveau de la strate que sera effectué l'échantillonnage, l'unité statistique de base est appelée généralement "segment", on procède à un sondage systématique de carrés ou segments d'une superficie définie (environ 50 ha). Le taux de sondage est fixé préalablement, on complètera néanmoins un minimum de 30 segments par strate. Ces segments seront investigués sur le terrain, l'enquête déterminera les pourcentages des différentes occupations du sol dans le segment. Ces segments feront également l'objet d'une étude et cartographie précise par classification des données SPOT multispectrales (analyse texturale) et donneront donc une nouvelle estimation des pourcentages d'occupations du sol.

La méthode dite de l'estimateur de régression permettra d'accroître la précision de l'estimation. Elle consiste à utiliser parallèlement les données terrain et les données de la classification sur les segments afin de construire une droite de régression. Cette droite de régression permet de calculer le biais, sous-estimation ou surestimation, introduit par la classification et ainsi corriger l'évaluation initiale au niveau global. Si on améliore par ce redressement les données statistiques, la cartographie au niveau régional ne s'en trouve pas plus précise. A cet égard, une méthode devrait faire l'objet d'attentions (Defourny et al., 1987).

Le développement d'un processus d'analyse diachronique des données sur chaque zone (banque de données locales) permettra un suivi des superficies mises en cultures par espèces d'années en années, tout en réduisant le temps et le coût des traitements.

5 CONCLUSIONS ET PERSPECTIVES

L'étude entreprise a permis de prouver qu'il était possible de différencier les surfaces de culture à l'aide des données SPOT multispectrales.

Cependant, l'analyse de la production d'une culture passe par l'estimation du produit surface - rendement.

Si la composante surface est obtenue par les enregistrements à haute résolution SPOT multispectral, l'estimation des rendements fait appel à d'autres critères. Elle nécessite, en effet, des observations fréquentes de l'état de la végétation tout au long de la saison agricole afin de permettre par l'intégration des paramètres une meilleure évaluation des rendements.

Le schéma global d'une étude des productions agricoles passe donc par l'intégration et l'interaction des différentes plate-formes satellitaires.

- Enregistrement LANDSAT MSS pour la délimitation et l'évaluation des superficies totales cultivées, de même que pour la stratification préalable au sondage.
- Enregistrement SPOT multispectral pour la détermination des surfaces cultivées par espèces au niveau local, classification des segments.
- Enquête de terrain qui vise à collecter les données (type de culture, pourcentages d'occupation,...) sur les segments spatiaux définis sur images télédéteçtées afin de corriger le biais des classifications régionales.

6 REMERCIEMENTS

Cette recherche a été effectuée dans le cadre du Programme National Belge R-D "Etude des ressources terrestres et marines par satellite" (Service du Premier Ministre - Programmation de la Politique Scientifique).

7 BIBLIOGRAPHIE

Bartholome, E., "Estimation des rendements agricoles à l'aide des indices de végétation". Séminaire "Amélioration des procédures d'enquêtes agricoles par télédétection" Centre AGRHYMET, Niamey, 1987, 31 p.

Defourny P., Massart M., Soyer J., "Méthodes de discrimination des champs traditionnels en milieu tropical humide (Shaba, Zaïre) par traitement numérique des données SPOT". Actes du colloque SPOT : SPOT 1 Utilisation des images, bilan, résultats, 1987 pp 399-406

Etude FAO, "Estimation des superficies cultivées et des rendements dans les statistiques agricoles", 1982.

Haralick, R.M., Shanmugan, K., Dinstein, I., "Textural features for image classification" IEEE Trans. Syst., Man, Cybern. Vol. SMC-3, 1973, pp. 610-621.

Lambin, E., "Apport de la télédétection satellitaire pour l'étude des systèmes agraires et la gestion des terroirs en Afrique occidentale. Exemples au Burkina Faso" Dissertation présentée en vue de l'obtention du grade de Docteur en sciences, Louvain la Neuve, 1988.

Meyer-Roux, J., "Estimation des surfaces cultivées à partir des données LANDSAT et des enquêtes de terrain" 1981, U.S.D.A. Statistical Research Division.

Meyer-Roux, J., Fournier, Ph., "Agricultural statistics and remote sensing french research and studies with LANDSAT MSS and SPOT simulations", 1982.

Tokerud, R.E., Quirein, J.A., "An assessment of local and related methodologies for conducting crop inventories". In : Remote sensing applications in developing countries, Remote Sensing Society, 1978, pp 91-101.

APPLICATIONS OF STEREO SPOT FOR MAPPING

J-P MULLER AND I.J.DOWMAN

DEPARTMENT OF PHOTOGRAMMETRY AND SURVEYING
UNIVERSITY COLLEGE LONDON
GOWER STREET,
LONDON WC1E 6BT
UNITED KINGDOM

ARPANET: jpmuller@cs.ucl.ac.uk

Telephone: (44)-1-380-7227. Fax: (44)-1-380-0453. Telex: 296273 UCLENG

ABSTRACT

The geometry of overlapping oblique SPOT images can be exploited to produce digital elevation models, orthoimages and line maps with a high degree of automation. An end-to-end system has been developed at University College London as part of an Alvey project in collaboration with Thorn EMI Central Research Laboratories, Laser-Scan Laboratories, RSRE Malvern and the Department of Computer Science, UCL. Funding was provided by SERC, the UK Department of Trade and Industry and the UK Ministry of Defence. The system generates mapping products using automatic matching of stereoscopic images to produce image disparities; a dynamic camera model which makes use of the SPOT header data with as few as 2 control points and which can be extended over a strip of models; intersection of points in the model space to produce a digital elevation model from the disparities and the resampling of the data using the DEM to produce orthoimages. Validation takes place through in-built checks and comparison with three-dimensional co-ordinates obtained by other methods such as aerial photographs or ground survey. Line information can at present be extracted from hard copy images on a Kern DSR analytical plotter but developments are taking place so that this can be done on the Sun workstation on which the digital system is running. Very fast processing speeds can be achieved through the use of a Parsys™ Supernode transputer machine which allows matching of 2 complete SPOT scenes in 120 minutes.

The current camera model gives an accuracy of better than 10m in plan and height when checked with 40 ground control points. An improved model indicates similar or better accuracy with only 2 ground control points. The digital terrain model has been checked against data derived from aerial photographs and gives a root mean square error of 11m when compared with 28053 points over a 12.4km x 6.9km area.

RADIOMETRIC COMPARISON OF LANDSAT-5 TM AND SPOT HRV 1 SENSORS FOR THE USE IN MULTIPLE SENSOR APPROACHES

J. Hill D. Aifadopoulos

Institute for Remote Sensing Applications
Lab. for Image Processing

Commission of the European Communities
Joint Research Center
I-21020 Ispra (Va), Italy

Abstract

This paper analyses the geometric and, in particular, the radiometric accuracy of SPOT HRV-1 data in comparison to Landsat-5 TM imagery. We use concurrent TM and SPOT images recorded almost simultaneously over the southern Ardeche region (France), both data sets being available on system corrected CCT. The TM and SPOT scenes could be registered to map projection with sub-pixel accuracy, while a common pixel size of 30 x 30 m was maintained. The mountainous character of the study region, however, required the use of digital elevation data for the compensation of relief induced distortions. The radiometric comparison of TM and SPOT uses the apparent reflectances of 28 reference targets representing a wide range of vegetated and non-vegetated cover types. Terrain slopes were generally less than 10 percent and considered insignificant. While a very good match of TM and SPOT reflectances in the visible bands could be stated, a considerable discrepancy was found concerning the calibration of the near-infrared channel of the SPOT HRV-1 sensor. A calibration adjustment was performed with reference to the contemporary TM scene, and both scenes were corrected for atmospheric absorption, scattering and pixel adjacency effects. It is demonstrated, that identical values of the normalized difference vegetation index (NDVI) can be obtained from both sensors, once proper calibration data are provided.

Key Words: Landsat-5 TM, SPOT HRV-1, geometric registration, radiometric comparison, sensor calibration, atmospheric corrections, NDVI.

1. Introduction

Multi-temporal approaches to monitoring, classification and mapping of agricultural crops and natural vegetation in Central Europe are often hampered by the lack of suitable imagery due to the frequent cloud cover. The joint use of Landsat-5 Thematic Mapper and SPOT data in multiple sensor approaches can help solving some acquisition problems. The processing of multi-temporal data involves techniques for data compression and image classification. In simple approaches, the classification of multiple sensor data sets would be based on all available data channels. A reduction of the data volume is mostly achieved by channel ratioing or the use of linear transformations. Any quantitative analysis of the resulting features, however, is facilitated if comparable wavebands can be used for the calculation of greenness indices (i.e. NDVI). This requires that the differences between comparable wavebands are not too great to permit direct comparisons.

Before, however, any effort is made to use multiple sensor configurations, the relationship between the sensor systems (spatial resolution, viewing geometry, orbital characteristics) and their corresponding wavebands needs to be evaluated.

2. Data sets

This paper analyses the geometric and, in particular, radiometric accuracy of Landsat-5 TM and SPOT HRV 1 data by using scenes which both were acquired at the same day (29-09-1986) over the JRC study site in the Departement Ardeche (France). The area is located about 40 km south of Lyon and extends from the Rhone valley (100 m amsl) towards the eastern 'Massif Central' where altitudes of about 1700 m are reached (Hill & Megier, 1988). The TM and SPOT scenes were recorded with a time difference of one hour which results in a minor change of illumination geometry (table 1), but it can be assumed that there was no significant change in atmospheric conditions. The SPOT data set includes two subsequent scenes from the same orbit

Table 1. TM and SPOT scene characteristics

| | TM | SPOT |
|------------------|------------------|------------------|
| Acquisition date | 29-Sep-1986 | |
| WRS frames | 197/29 | 47/260&231 |
| Scene location | 44.6 N 3.72 E | 44.2 N 3.85 E |
| Acquisition time | 9.49 | 10.47 |
| Sun elevation | 37.42 | 42.20 |
| Sun azimuth | 145.92 | 163.95 |
| Instr. tilt | — | R2.4 |

which are both located within the coverage of TM quarters 2 and 4. Since the observation geometry for the two data sets comprise only view angles of 0 to 5 degree for TM, and -4.1 to -2.0 degree for SPOT, nadir viewing is assumed for the radiometric analysis. The TM image was acquired as system corrected CCT, the SPOT scene is a level 1B product.

3. Geometric registration

The first step in generating multiple sensor data sets is the geometric registration of the images to a common map grid. In this case, a pixel size of 30 x 30 m was maintained for the geocoded data, which implied the resampling of SPOT data to TM ground resolution.

It is known that system corrected TM data and SPOT level 1b products are of exceptionally good geometric quality which allows, over level terrain, geodetic rectifications to sub-pixel precision by means of first order polynomials. But for high resolution satellite imagery such scene distortions have become important which are due to local variations in terrain altitude. The resulting horizontal displacement of ground features (relief displacement) is known to cause misregistration of multi-temporal overlays, especially in high relief areas as our study region.

Since TM system corrected data and SPOT level 1B products are already corrected for internal scan distortions (i.e. non-linearity of mirror motion, earth curvature effects, panoramic distortion), scan skew and sensor attitude variations, both images can be assumed to form geometrically a series of near-orthogonal projections in along-track direction and, across-track, a series of parallel perspective projections (Kohl & Hill, 1988).

The elimination of relief displacement effects can therefore be limited to a correction of the respective pixel position within a scan line

$$P_y' = P_y \pm h \tan \theta \quad (3.1)$$

where P_y' denotes the corrected y-coordinate, and P_y is the original position of the pixel within the image line; h is the terrain elevation at the respective map position and θ gives the observation angle at the image pixel which, in the case of SPOT, includes also the instrument tilt angle ω . The sign of the offset depends on the pixel position relative to the line nadir (Kohl & Hill, 1988).

The geometric registration process of both scenes consequently combined GCP-based affine transformations and the correction of terrain induced image distortions. By using a commercially available DEM of 250 x 250 m grid-resolution only, sub-pixel accuracy was achieved for the geometric rectification of both data sets.

4. Radiometric comparison and sensor calibration

Many research and operational applications require the conversion of digital count values (DN) to quantitative physical values. The apparent at-satellite reflectance ρ^* compensates for differences in illumination geometry and, assuming lambertian ground reflectance, can be computed for a given spectral band with

$$\rho^* = \frac{\pi L}{E_0 \cos(\theta_0) d} \quad (4.1)$$

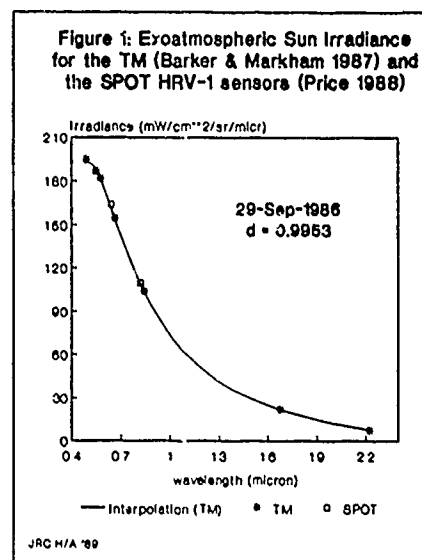
where θ_0 is the solar zenith angle and d is a correction coefficient for the actual sun-earth distance. The exoatmospheric solar irradiance E_0 for the Thematic Mapper and SPOT HRV-1 bandpasses is obtained from Markham & Barker (1987) and Price (1988) (figure 1). The computation of the at-satellite measured spectral radiance involves the use of respective calibration functions.

4.1 TM calibration

Thematic Mapper radiances L_{tm} [$\text{mW}/\text{cm}^2/\text{sr}/\mu\text{m}$] for the six reflective channels are calculated with

$$L_{tm} = a_0 + a_1 \text{DN} \quad (4.2)$$

In-flight calibration gains (a_1) and offsets (a_0) have been assessed through a series of contemporary ground measurements at the large



uniform gypsum sand area of White Sands, New Mexico (Slater et al., 1986, 1987). Based on these measurements, updated calibration constants were defined for TM bands 1 to 4; no changes were applied to TM band 5 and 7 (Hill & Sturm, 1989) (table 2). The validity of the new calibration functions was confirmed by a number of ground measure-

Table 2 Updated TM calibration constants acc. Slater et al. (1986)

| Band | pre-flight (ESA) | | Update 1984/1985 | |
|------|------------------|--------|------------------|-----------|
| | a_0 | a_1 | a_0 | a_1 |
| TM 1 | -0.1009 | 0.0636 | -0.1331 | 0.0727 |
| TM 2 | -0.1919 | 0.1262 | -0.2346 | 0.1385 |
| TM 3 | -0.1682 | 0.0970 | -0.1897 | 0.1102 |
| TM 4 | -0.1819 | 0.0914 | -0.1942 | 0.0885 |
| TM 5 | -0.0398 | 0.0126 | unchanged | unchanged |
| TM 7 | -0.0203 | 0.0067 | unchanged | unchanged |

ments which were carried out by JRC's radiometric measurement group during TM overflights in 1984 (Maracci et al., 1986), 1986 and 1988 (Maracci, pers. communication). Since no evidence was found for a strong degradation of the TM system response with time, the updated TM calibration was chosen as reference for the following comparisons.

4.2 SPOT calibration

At-satellite radiance L_{spot} [$\text{W}/\text{m}^2/\text{sr}/\mu\text{m}$] for each SPOT HRV band is obtained from

$$L_{spot} = \frac{1}{\lambda} \text{DN} \quad (4.3)$$

The respective calibration factors a are monthly updated by using the on-board sun calibrator and provided on each CCT. The absolute calibration gains for the scene under study are given in table 3.

Table 3. SPOT calibration gains

| | |
|-----|---------|
| XS1 | 0.90482 |
| XS2 | 0.84505 |
| XS3 | 0.92705 |

Figure 2: Lac Coucouron (Ardeche)
Comparison of the apparent reflectance
from coincident TM and SPOT imagery

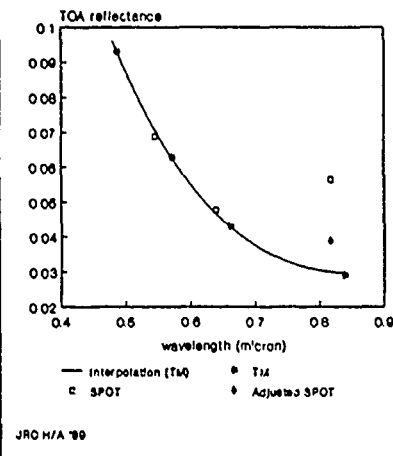


Figure 3: River Bank, Vallon (Ardeche)
Comparison of the apparent reflectance
from coincident TM and SPOT imagery

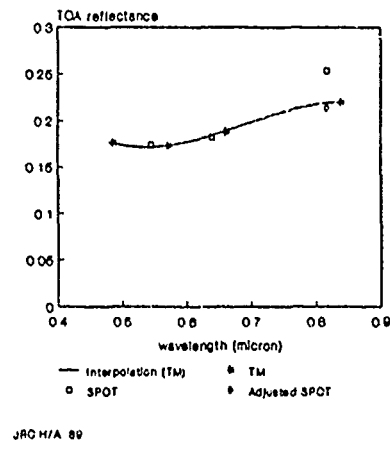
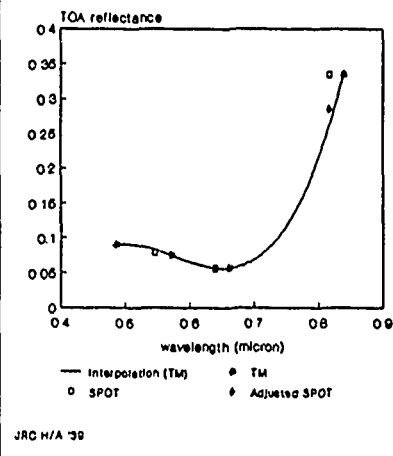


Figure 4: Deciduous Forest (Ardeche)
Comparison of the apparent reflectance
from coincident TM and SPOT imagery



The in-flight calibration device of SPOT was also tested during a measurement campaign which took place at White Sands in March 1986. After respective adjustments, a precision of $\pm 5.4\%$ was specified for the SPOT HRV calibration (Begni et al., 1986).

4.3 Radiometric comparison of the TM and SPOT HRV-1 instrument

Totally 28 test sites were selected for the comparison between TM and SPOT radiance levels, 15 of which are vegetated targets with a good variety of canopy conditions, 8 are bare soil surfaces and 5 are water targets. All sites are located on flat terrain in order to minimize effects which are due to the different sun illumination.

A comparison of the apparent reflectances recorded from TM and SPOT revealed always a very good agreement between the comparable bands in the visible part of the spectrum. This proves the validity of the respective calibration functions and, in addition, it confirms the assumption of stable atmospheric conditions between the data acquisitions. For all targets, however, too high a signal level was found in the near-infrared channel of SPOT, which indicates that the 1986 calibration update for the HRV-1 band XS3 has not been adequate (figure 2-4).

Since there is strong evidence that, in 1986, the updated TM calibration yields reliable at-satellite radiances, it was attempted to correct the XS3 calibration with reference to the contemporary TM measurements. The corrected ρ_{XS3}^* for each reference target has been approximated by using a polynomial interpolation between the TM bands 1 to 4. Interpolated SPOT XS3 radiances L_{XS3}^* were then obtained by inverting equation (4.1) (figure 2-4).

A linear regression between interpolated radiances and XS3 grey levels gave the updated calibration function

$$L_{XS3}^* = -0.269 + 0.0962 DN_{XS3} \quad (4.4)$$

with a regression coefficient of 0.9978. The results obtained with this calibration agree clearly better to the coincident TM measurements (figure 2-4), and have therefore been used in the further analysis.

Since the SPOT instrument by definition has no calibration offset, it is assumed that the bias value must be attributed to remaining uncertainties in the TM calibration or the interpolation method. The multiplicative term only would correspond to an absolute calibration

gain $a = 1.0395$. Compared to the coefficient provided on tape, this means an increase of 12.13 %, which conforms rather well to the last communications about the absolute calibration of SPOT. New measurements taken over White Sands and snow surfaces have actually confirmed the calibration of bands 1 and 2, but the calibration coefficient for band 3 had to be increased by 8 % (Begni 1988).

5. Comparison of TM and SPOT vegetation indices

Field experiments have shown that important parameters as leaf area index, absorbed photosynthetic active radiation and wet and/or dry biomass are related to the normalized difference vegetation index (NDVI) or similar combinations of wavebands in the visible and near-infrared. It is evident, that the joint use of TM and SPOT data for the quantitative observation of such variables depends heavily on the reliability of each systems's calibration.

Since atmospheric conditions may introduce additional variability in the vegetation indices obtained from different sensors, the TM and SPOT scenes were corrected for atmospheric absorption, scattering and pixel adjacency effects according to a radiative transfer model derived from the one developed by Tanre et al. (1979):

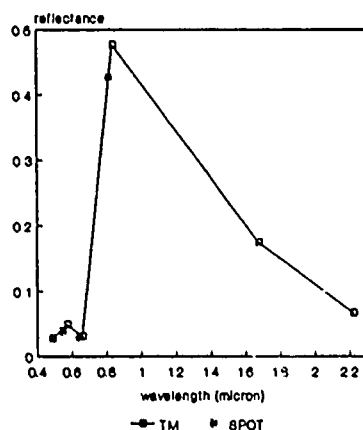
$$\rho^* = t_{03} \left\{ \rho_{at} + \frac{T(\mu_0) [t_d(\mu) \rho_t + t_s(\mu) \langle \rho \rangle]}{1 - \langle \rho \rangle s} \right\} \quad (5.1)$$

where ρ_t is the target reflectance, $\langle \rho \rangle$ the background contribution to the apparent reflectance and s is the spherical albedo. ρ_{at} denotes the intrinsic atmospheric signal component, $T(\mu_0)$ the total downward, $t_d(\mu)$ the diffuse and $t_s(\mu)$ the scattered upward transmittance; t_{03} gives the ozone transmittance.

The atmospheric key parameter of aerosol optical thickness τ was estimated from the TM signal obtained over clear lakes, and then fitted to the so-called Angstrom relation which allowed to retrieve compatible processing parameters for the SPOT scene.

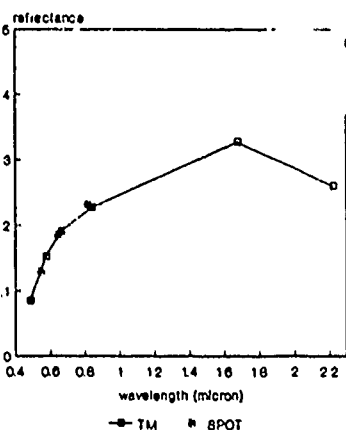
For each of the reference targets, SPOT and TM vegetation indices were calculated based on the retrieved ground reflectance ρ_t (figure 5-6) in the corresponding channels of the visible red and the near-infrared. The comparison was based on vegetated and bare soil targets only, and it

Figure 5: Maize, Beaulieu (Ardeche)
Target reflectance from TM and SPOT



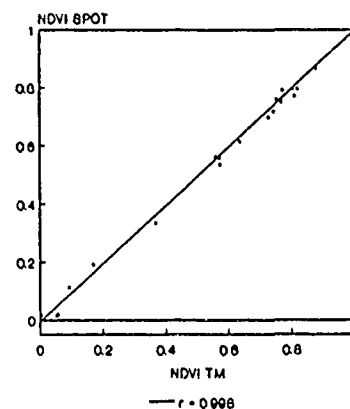
JRO H/A '88

Figure 6: Bare Soil, Beaulieu (Ardeche)
Target reflectance from TM and SPOT



JRO H/A '88

Figure 7: Relation between the NDVI
from atmospherically corrected
TM and SPOT imagery (29-09-1988)



JRO H/A '88

revealed a significant linear relation ($r = 0.998$) between the NDVI of SPOT and TM (figure 7):

$$NDVI_{spot} = -0.0054 + 0.9904 NDVI_{tm} \quad (5.2)$$

It is demonstrated that practically identical index values are obtained, once suitable calibration data are provided. These results completely agree with existing evaluations of vegetation indices for simulated Landsat-5 TM and SPOT-1 data (Gallo & Daughtry, 1987).

6. Conclusions

TM and SPOT data can be registered to a common map grid of 30 x 30 m ground resolution with sub-pixel accuracy. In case of mountainous relief, digital elevation data must be included in the correction. Local misregistrations are only due to the different resampling algorithms used throughout the system corrections, since TM products from EURIMAGE are not available with cubic convolution resampling.

The analysis of coincident TM and SPOT acquisitions over the same area gave no evidence for a further calibration adjustment in the visible bands of both systems. A strong discrepancy, however, was found between TM band 4 and the SPOT near-infrared channel XS3, which could be attributed to a miscalibration of the SPOT band. In the meanwhile, SPOT IMAGE has confirmed that the calibration of XS3 had to be adjusted, but the date of the reinitialization of the on-board calibrator was not communicated.

The validity of the TM calibration updates, which were based on the ground measurements at White Sands, has been confirmed. This suggests that TM can be used as calibration reference for further evaluations of this type.

TM and SPOT data permit a quantitative analysis in terms of ground reflectances, once suitable calibration data are provided. Since both systems provide identical greenness indices (NDVI), a wide range of applications in agricultural monitoring and vegetation observation can be approached by using multiple sensor data sets which involve TM and SPOT imagery.

7. References

- Begni, G., M.C. Dingirard, R.D. Jackson & P.N. Slater, "Absolute calibration of the SPOT-1 HRV Cameras", SPIE vol. 660, 66-76, 1986.
- Begni, G., "Absolute calibration of Spot data", Spot Newsletter, no. 10, 2-3, 1988.
- Gallo, K.P. & C.S.T. Daughtry, "Differences in vegetation indices for simulated Landsat-5 MSS and TM, NOAA-9 AVHRR, and SPOT-1 sensor systems", Remote Sensing of Environment, 23, 439-452, 1987.
- Hill, J. & J. Megier, "Regional land cover and agricultural area statistics and mapping in the Departement Ardeche, France, by use of Thematic Mapper data", Int. J. Remote Sensing, vol. 9, no. 9/10, 1573-1585, 1988.
- Hill, J. & B. Sturm, "Radiometric corrections of multi-temporal Thematic Mapper data for the use in land-cover classification and vegetation monitoring", Int. J. Remote Sensing, acc. for publication, 1989.
- Kohl, H.G. & J. Hill, "Geometric registration of multi-temporal TM data over mountainous areas by use of a low resolution digital elevation model", Proc. 8th EARSeL Symp. on Alpine and Mediterranean Areas: A Challenge for Remote Sensing, Capri (Naples), Italy, 17-20 May 1988, 323-335, 1988.
- Maracci, G.C., B. Hoogood & G. Andreoli, "Measurements of spectral signatures in less favoured areas (LFA): a contribution to the definition of a remote sensing multi-temporal experiment", ESA/EARSeL Symp. on Europe from Space, Lyngby, DK, 25-28 June 1986, ESA SP-258, 215-218, 1986.
- Markham, B.L. & J.L. Barker, "Thematic Mapper bandpass solar exoatmospheric irradiances", Int. J. Remote Sensing, vol. 8, no. 3, 517-523, 1987.
- Price, J.C., "An Update on visible and near infrared calibration of satellite instruments", Remote Sensing of Environment, 24, 419-422, 1988.
- Slater, P.N., S.F. Diggar, R.G. Holm, R.D. Jackson, Y. Mao, M.S. Moran, M. Palmer & B. Yuan, "Absolute radiometric calibration of the Thematic Mapper", SPIE, vol. 660, 2-8, 1986.
- Tanre, D., M. Herman, P.Y. Deschamps & A. de Lefre, "Atmospheric modelling of the background contribution upon space measurements of ground reflectance, including bi-directional properties", Appl. Optics, 18, 3587-3594, 1979.

NEURAL NETWORK APPROACHES VERSUS STATISTICAL METHODS IN CLASSIFICATION OF MULTISOURCE REMOTE SENSING DATA

J.A. Benediktsson, P.H. Swain and O.K. Ersoy
School of Electrical Engineering

and
Laboratory for Applications of Remote Sensing
Purdue University

W. Lafayette, IN 47907, U.S.A.
Telephone: (317) 494 - 1743, FAX: (317) 494 - 6440

ABSTRACT

Neural network learning procedures are applied to the classification of multisource remote sensing data. Statistical methods are used to classify the same data. Experimental results are given and a comparison is made between the two different approaches. The main emphasis in the comparison is in terms of classification accuracy but other factors such as ease of implementation and speed of algorithms are also considered. Two methods show very good performance: statistical multisource analysis and in the neural network case, the generalized delta rule.

Keywords: Classification, multisource data, three-layer neural networks, statistical distributions.

1. INTRODUCTION

Recently, there has been a great resurgence of research in neural networks. New and improved neural network models have been proposed to classify complex data. In the remote sensing community, the question of how well these neural network models work as classifiers, compared to statistical classification methods, is very important. In this paper neural networks for classification of multisource remotely sensed and other geographic data are implemented and compared to some statistical approaches using preliminary experimental results. By multisource data we mean data from more than one distinct data source. In the experiments the data are Landsat MSS data (4 spectral bands) and topographic data which consist of elevation (1 data channel), slope (1 data channel) and aspect (1 data channel). The area used for classification, a mountainous area in Colorado, is the same area used in [1]. The training data and test data are also the same as used in [1], 831 training pixels and 1188 test pixels. The area has 10 ground cover classes, most of them forest types as shown in table 1. It is very difficult to distinguish among these forest types using just the Landsat MSS data alone; they show very similar spectral response. With the help of elevation, slope and aspect, the various forest types can be distinguished. A problem with using conventional multivariate statistical approaches for the topographic data is that the Gaussian distribution cannot be assumed for the ground cover classes as in the case of the Landsat data. This means that the conventional Gaussian maximum likelihood method is not appropriate for application to topographic data. Another common problem with statistical classification methods is that the data sources are not equally reliable. This means that the data sources need to be weighted according to their reliability but most statistical classification methods do not

have such a mechanism.

The neural networks, on the other hand, are non-parametric in the sense that no prior knowledge of the statistical distribution of the data is needed. This is an obvious advantage over most statistical classification methods, which face the problem of modeling the data, a difficult task when the distribution functions are not known or when the data are distinctly non-Gaussian. The neural networks also automatically take care of the problem involving how much weight each data source should have.

2. PRELIMINARY EXPERIMENTAL RESULTS USING STATISTICAL APPROACHES

Four statistical methods were used in the preliminary experiments performed here: 1) minimum Euclidean distance [2], 2) maximum likelihood algorithm for Gaussian data [2], 3) minimum Mahalanobis distance for Gaussian data [2] and 4) statistical multisource classification [1,3]. The first three methods are "simple" stacked-vector approaches which are not expected to give a good performance in terms of classification accuracy for the multisource data. Our initial attempt at classifying the data with the first three methods was to use the training fields for training and the test fields for testing. Using the training fields directly for training was not possible because methods 2 and 3 use the inverse of the class specific covariance matrices in the classification. Some of the information classes have no variation in the elevation, slope and aspect data. Therefore, the covariance matrices become singular for those classes. To be nevertheless able to apply methods 2 and 3 in classification, the data was clustered, with the initial means being selected as the means of the training classes. The clustering algorithm used (ISODATA algorithm) converged in 13 iterations and gave 10 very separable clusters. The mean vectors and the covariance matrices of these 10 clusters were then used in the classifications with algorithms 1, 2 and 3. The results of the classifications are shown in tables 2.a, and 2.b.

Although the data, which we know are not Gaussian, were assumed to be Gaussian, in application of the maximum likelihood algorithm and the Mahalanobis distance, these algorithms showed better results than the Euclidean distance. However, the overall accuracy of the test pixels was only about 49.5% for both the maximum likelihood method and the Mahalanobis distance.

As noted above, the multisource data set, the training data and test data are the same as used in [1] with statistical multisource analysis. The statistical multisource algorithm treats the data sources independently in training and combines the information from the sources in classification. In [1] the Landsat data are treated as Gaussian but all the data in the topographic data sources are non-Gaussian. If the

* This research is supported by the National Aeronautics and Space Administration Contract No. NAGW-925

best results reported in [1] (i.e., with the weighting that gives the highest accuracy for training data and the weighting that gives highest accuracy for testing) and displayed in tables 3.a and 3.b are compared to the results shown in tables 2.a and 2.b, it is clear that the statistical multisource algorithm is superior to all the "simple" statistical methods. Also, the "low" classification accuracy reported in [1] shows how difficult it is to classify the data correctly.

3. THE NEURAL NETWORK APPROACH

A neural network is a network of neurons where a neuron can be described in the following way: A neuron has many input signals (continuous-valued) $x_j, j = 1, 2, \dots, N$ which represent the activity on the current input line or the momentary frequency of neural impulses delivered by another neuron to this input [5]. In the simplest formal model of a neuron, the output frequency o is often approximated by a function

$$o = K \phi \left(\sum_{j=1}^N w_{ji} x_j - \theta \right)$$

where K is a constant and ϕ is a nonlinear function which takes the value 1 for positive arguments and -1 (or 0) for negative arguments. The w_j are called *synaptic efficacies* [5] or weights, and θ is a threshold.

In the neural network approach to pattern recognition the neural network operates as a black box which receives a set of input vectors x (observed signals), and produces responses o_i on its output ports i ($i = 1, \dots, L$ where L depends on the number of information classes). A general idea followed in neural network theory is that the outputs are either $o_i = 1$ (if neuron i is active for the current input vector x) or $o_i = -1$ (if it is inactive). This means the signal values are coded as binary vectors and for a specific input vector x the outputs give the binary representation of its class number. The process is then to learn the weights through an iterative training procedure. When the training procedure has converged, the data are fed into the network for the classification and the network returns at the output the binary representation of the class number of each pixel.

3.1 Preliminary Experimental Results Using Neural Network Models

Two neural network approaches were implemented in experiments to classify the data: the delta rule [4] and the generalized delta rule [4]. The neural networks were trained with binary input vectors. Since five of the seven data channels take values in the range from 0 to 255, each data channel was represented by 8 bits and therefore 8 input neurons. The total number of input neurons was $7 \times 8 = 56$. Since the number of training classes was 10, the number of output neurons was selected as 4. The weights in the neural networks were said to converge if the cumulative squared difference of weights for each connection between successive passes through the training data was less than 0.1.

3.1.1 Experiments with the Delta Rule

The delta rule for updating weights in presentation of input pattern number k can be written as:

$$W(k) = W(k-1) + \eta[t(k) - W(k-1)x(k)]x^T(k)$$

where $x(k)$ is the input pattern presented on the k th presentation, $t(k)$ the desired output for presentation k , $W(k)$ the state of the weight matrix describing the network after k presentations and η a gain factor (learning rate). A possible choice of η is $\eta = C/\text{ncycle}$, where C is a constant and ncycle the number of the learning cycle. That particular choice of η forces the weight matrix $W(k)$ to converge after several iterations. In all the experiments below, the initial gain factor was arbitrarily selected as 0.3. The delta rule which is identical to the mathematical method of stochastic approximation for regression problems cannot be used to

discriminate data that are not linearly separable and fails for instance in the learning of a XOR function.

Since this rule does not even guarantee a solution in the linearly separable case, it was not expected to perform well on this rather difficult data set. Using the gain factor as 0.3 for all cycles, it did not converge. By using a decaying gain factor: $0.3/\text{ncycle}$ the algorithm converged in 68 cycles. The results using the delta rule are not acceptable (see A in tables 4.a and 4.b).

3.1.2 Experiments with the Generalized Delta Rule

The application of the generalized delta rule or the principle of back propagation of errors involves two phases. During the first phase the input is presented and propagated forward through the network to compute the output value o_{pj} in presentation p for each unit j , i.e.,

$$o_{pj} = f_j(\text{net}_{pj})$$

where $\text{net}_{pj} = \sum_i w_{ji} o_{pi}$, w_{ji} is the weight of the connection

from unit i to unit j and f_j is the semilinear activation function at unit j which is differentiable and nondecreasing. A widely used choice for a semilinear activation function is the sigmoid function which is used in our experiments:

$$f_j(x) = 1/(1 + e^{-(x + \theta_j)})$$

where θ_j is a bias, similar to a threshold. The output is compared with the desired values, resulting in an error signal δ_{pj} for each output unit.

The second phase involves a backward pass through the network (analogous to the initial forward pass) during which the error signal is passed to each unit in the network and the appropriate weight changes are made according to:

$$\Delta_p w_{ij} = \eta \delta_{pj} o_{pi}$$

This second, backward pass allows the recursive computation of δ [4].

The generalized delta rule was implemented in the experiments with three layers (input, output and hidden layers) where the hidden layer had 32 units, an arbitrarily selected number. Using the gain factor as $0.3/\text{ncycle}$, the generalized delta rule converged very fast, i.e., in 6 cycles. The classification results for this experiment are marked B in tables 4.a and 4.b. Although these results were much better than the ones obtained with the delta rule, we could expect even better results when a constant gain factor is used. The reason for this is that the decaying gain factor forces the weight matrix to converge prematurely as in the delta rule above.

Using a constant gain factor (0.3) the generalized delta rule converged extremely slowly. The cumulative squared difference between the weights oscillated up and down, but decayed very slowly. After 200 cycles the cumulative squared difference between the weights was 78.43. The classification results using the constant gain factor are much better for the training data in terms of accuracy than when the decaying gain factor was used. These results are marked C in tables 4.a and 4.b.

The generalized delta rule with the constant gain factor converged in 344 cycles. The classification results for this case are by far the best for the training data (91.9%), as shown in table 4.a (D). Compared to the other neural network experiments the results are also best for the test data (51.5%) as shown in table 4.b (D). The classification accuracy of the training data is very satisfactory since the data set is, as said before, very hard to classify accurately. On the other hand the learning algorithm is computationally very complex and took 22 hours to converge on a Gould NP-1 machine.

Although the generalized delta rule is superior to the other methods in classification of training data it does not do nearly as well in classifying the test data. This shows how

important it is to select representative training samples when training a neural network. However, the training data used here, as in [1], might be questionable since only one training field was selected for each information class. This implies that each information class only had one subclass! To see how the generalized delta rule would do if approximately 60% of each field were used for training and the other 40% for testing, another experiment was conducted. In this experiment the number of training samples was 1307 and the number of test samples was 712. Using this training data the generalized delta rule with a constant gain factor did not converge in more than 40 hours on a Gould NP-1 machine. Since the machine was going down rather often at the time the experiments were done, this particular experiment could not be finished. On the other hand using the decaying gain factor with the new training data the generalized delta rule converged in 6 cycles as before. The results from this experiment are shown in tables 5.a and 5.b.

In tables 5.a and 5.b we see the best overall classification accuracy achieved for the test pixels (66.6%). That accuracy is a little bit higher than for the training pixels (64.7%). An accuracy of 66.6% for the test pixels is very good for this data set. We can expect the results with the generalized delta rule with a constant gain factor to be even better. But the results in tables 5.a and 5.b illustrate how important it is to have representative training samples to get good results.

4. Discussion

Of the neural network models tried, the delta rule as expected did not perform well. On the other hand the generalized delta rule showed great potential as a pattern recognition method for multisource remotely sensed data. It is superior to the statistical methods used in terms of classification accuracy of training data. It has the advantage that it is non-parametric in the sense that we do not have to know anything about the statistical distribution of the data. This is an obvious advantage over most statistical methods, which for instance face the problem of modeling the data, which is difficult when we have no knowledge of the distribution functions or when the data are non-Gaussian. It also eliminates the problem of determining how much influence a source should have in the classification. That is still a problem for statistical methods, e.g., statistical multisource analysis as discussed in [1].

However, the generalized delta rule is computationally complex. When the sample size is large the learning time can be very long. In this respect, the experiments show how important our selection of the gain factor is. Selection of a constant gain factor gave the best classification results. A constant gain factor, η , increases the weight changes in proportion to η and the optimum gain factor is the one which has the largest value that does not lead to oscillation. It is very difficult to determine this optimum value. In the experiments the value 0.3 was arbitrarily chosen. That value is not large but anyhow led to oscillation in some cases. We said the training procedure converged when the weight matrices "did not" change between learning cycles. With this definition of convergence, the training process of the neural networks sometimes shows convergence problems when a constant gain factor is selected. A decaying gain factor can be a better choice, to get relatively good results fast.

To perform well the neural network models have to be trained by representative training samples. If that can be achieved, our results show that a three-layer net can outperform the statistical methods.

As reported in [1] the statistical multisource classification algorithm worked well for combination of multispectral and topographic data and showed very good accuracy in classifying the data. The statistical multisource classification algorithm is not as sensitive to representative training samples as the neural network models and outperformed them in classifying test data. This algorithm does not have the same convergence problems as the neural

network models but on the other hand requires more preparations from the analyst. It also has to be pointed out that only relatively "simple" methods were used in [1] to control the influence (weight) of the sources and model the non-Gaussian data. When the weighting is done more properly, we can expect the statistical multisource algorithm to perform even better. The other statistical classification algorithms (Euclidean distance, Mahalanobis distance and the maximum likelihood algorithm for Gaussian data) did not show good performance, but that was expected.

The main advantage statistical classification algorithms have over the neural network models is in general that if we know the distribution functions of the information classes these methods can be simple, fast and accurate. In many cases, as for instance in multisource classification, we cannot assume we know the distribution functions. Therefore, three-layer neural network models can be more appropriate, in particular if the training process converges in a reasonable amount of time. Faster learning methods are a subject of current research. We are also continuing to investigate the performance of both neural network models and statistical methods using additional data sets.

References

1. J.A. Benediktsson and P.H. Swain, "A Method of Statistical Multisource Classification with a Mechanism to Weight the Influence of the Data Sources," Proceedings of IGARSS '89, Vancouver, Canada, July 10-14th, 1989.
2. J.A. Richards, *Remote Sensing Digital Image Analysis - An Introduction*, Springer-Verlag, Berlin, W. Germany 1986.
3. T. Lee, J.A. Richards and P.H. Swain, "Probabilistic and Evidential Approaches for Multisource Data Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, no. 3, pp. 283-293, May 1987.
4. D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning Internal Representation by Error Propagation," *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1, D.E. Rumelhart and J.L. McClelland (eds.), Cambridge, MA: MIT Press, pp. 318-362, 1986.
5. T. Kohonen, "An Introduction to Neural Computing," *Neural Networks*, vol. 1, no. 1, pp. 3-16, 1988.

Table 1

| Class # | Information Class |
|---------|----------------------------------|
| 1 | water |
| 2 | colorado blue spruce |
| 3 | mountane/subalpine meadow |
| 4 | aspen |
| 5 | ponderosa pine |
| 6 | ponderosa pine/douglas fir |
| 7 | engelmann spruce |
| 8 | douglas fir/white fir |
| 9 | douglas fir/ponderosa pine/aspen |
| 10 | douglas fir/white fir/aspen |

Table 2. Classification of (a) training samples, (b) test samples after clustering. The following statistical methods were used:
 1) the minimum Euclidean distance (ED),
 2) the maximum likelihood method (ML) and
 3) the minimum Mahalanobis distance (MD).

Table 2.a

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|-----|----|----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| ED | 100 | 0 | 0 | 64 | 0 | 37 | 85 | 0 | 0 | 0 | 58.2 | |
| ML | 100 | 0 | 0 | 45 | 0 | 37 | 100 | 0 | 0 | 20 | 60.9 | |
| MD | 100 | 0 | 0 | 45 | 0 | 37 | 100 | 0 | 0 | 18 | 60.8 | |
| #: | 408 | 88 | 45 | 75 | 105 | 126 | 224 | 32 | 25 | 60 | 1188 | |

Table 2.b

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|----|----|----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| ED | 95 | 0 | 0 | 28 | 10 | 63 | 56 | 0 | 52 | 0 | 46.6 | |
| ML | 95 | 0 | 0 | 26 | 10 | 63 | 90 | 0 | 48 | 0 | 49.2 | |
| MD | 95 | 0 | 0 | 26 | 10 | 63 | 93 | 0 | 56 | 0 | 49.7 | |
| #: | 195 | 24 | 42 | 65 | 139 | 188 | 70 | 44 | 25 | 39 | 831 | |

Table 3. Classification of (a) training samples, (b) test samples using the statistical multisource classification algorithm. For A the weights assigned to the sources were 1. (Landsat MSS), .4 (elevation), .4 (slope) and .4 (aspect). For B the weights were 1., .8, .4, .6 (using the same order of the sources as above).

Table 3.a

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|-----|----|----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| A: | 100 | 100 | 11 | 71 | 33 | 73 | 95 | 0 | 0 | 58 | 78.0 | |
| B: | 100 | 99 | 0 | 64 | 37 | 15 | 97 | 38 | 0 | 24 | 77.8 | |
| #: | 408 | 88 | 45 | 75 | 105 | 126 | 224 | 32 | 25 | 60 | 1188 | |

Table 3.b

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|----|----|----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| A: | 86 | 0 | 0 | 52 | 49 | 76 | 97 | 0 | 0 | 0 | 57.9 | |
| B: | 100 | 0 | 0 | 51 | 38 | 84 | 97 | 0 | 0 | 0 | 60.8 | |
| #: | 195 | 24 | 42 | 65 | 139 | 188 | 70 | 44 | 25 | 39 | 831 | |

Table 4. Classification of (a) training samples, (b) test samples using the following neural network models: A: the delta rule (gain factor: 0.3/ncycle), B: the generalized delta rule (gain factor: 0.3/ncycle, hidden units: 32), C: the generalized delta rule (gain factor: 0.3, hidden units: 32, learning cycles: 200), D: the generalized delta rule (gain factor: 0.3, hidden units: 32, learning cycles: 344).

Table 4.a

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|-----|----|-----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| A: | 38 | 63 | 18 | 53 | 0 | 58 | 95 | 0 | 0 | 45 | 48.1 | |
| B: | 100 | 15 | 38 | 81 | 2 | 77 | 96 | 50 | 24 | 15 | 71.6 | |
| C: | 100 | 0 | 33 | 91 | 40 | 98 | 100 | 97 | 60 | 92 | 82.5 | |
| D: | 100 | 100 | 64 | 95 | 63 | 79 | 100 | 75 | 100 | 98 | 91.9 | |
| #: | 408 | 88 | 45 | 75 | 105 | 126 | 224 | 32 | 25 | 60 | 1188 | |

Table 4.b

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|-----|-----|-----|----|----|----|------|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| A: | 25 | 71 | 4 | 17 | 0 | 32 | 97 | 14 | 0 | 36 | 27.3 | |
| B: | 94 | 17 | 33 | 43 | 2 | 43 | 99 | 0 | 0 | 8 | 46.3 | |
| C: | 87 | 4 | 2 | 28 | 1 | 48 | 94 | 9 | 0 | 64 | 45.2 | |
| D: | 89 | 88 | 26 | 40 | 13 | 42 | 96 | 11 | 0 | 69 | 51.5 | |
| #: | 408 | 88 | 45 | 75 | 105 | 126 | 224 | 32 | 25 | 60 | 1188 | |

Table 5. Classification of (a) a new set of training samples, (b) a new set of test samples using the generalized delta rule with the gain factor 0.3/ncycle and 32 hidden units. Weights were determined after 6 learning cycles

Table 5.a

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|-----|-----|-----|-----|----|----|----|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| | | 100 | 0 | 0 | 84 | 16 | 76 | 90 | 0 | 0 | 67 | 64.7 |
| #: | 362 | 68 | 43 | 110 | 199 | 740 | 158 | 76 | 30 | 51 | 1307 | |

Table 5.b

| | | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|----|-----|--|----|----|----|----|-----|----|----|----|-----|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| | | 100 | 9 | 0 | 90 | 22 | 64 | 91 | 0 | 0 | 46 | 66.6 |
| #: | 241 | 44 | 44 | 30 | 45 | 74 | 136 | 30 | 20 | 48 | 712 | |

Statistical Methods for Cover Class Mapping using Remotely Sensed Data

N.A. Campbell and J.F. Wallace

Division of Mathematics and Statistics, CSIRO
Wembley 6014 W.A. AUSTRALIA

Tel: (619) 387 0294 Fax: (619) 387 6046

ABSTRACT

The paper discusses statistical considerations and developments in the use of remotely sensed spectral data for land cover mapping.

The initial stage is to produce colour composite displays of the data. From these and available ground information, training classes are selected.

Canonical variate analysis is used to provide information on the clustering and separation of the training classes in the multispectral data space. The definition and selection of spectral classes is followed by an allocation procedure. Pixels are assigned probabilities of class membership according to their relative and absolute closeness to the spectral classes.

The whole procedure allows for iterative refinement. In particular, class labels can be used to improve the classification procedure.

Keywords: canonical variate analysis; classification; posterior probabilities; typicality index.

INTRODUCTION

The CSIRO Division of Mathematics and Statistics has a research project on the Analysis of Remotely Sensed Data. The general objectives of the Project are to assess, develop and implement methods for the analysis of remotely sensed data

and to collaborate in research projects with other CSIRO Divisions and appropriate Government and industry groups.

The collaborative studies have generally been concerned with the mapping of pixels into various land cover classes, in both agricultural and rangeland settings. Assigning pixels to one or other of a number of reference classes, such as crop, pasture and bush in agricultural problems, and/or determining the proportions of cover types within a pixel, such as bare soil and shrub in rangeland problems, are common to most studies.

A general statistical approach to cover class mapping in agricultural settings has evolved. Spectral plots of the pixels are used to assist in identifying bands for an initial three-colour display of the data. Analyses which lead to linear combinations of the spectral bands can also be used to form the spectral plots. Training classes consisting of pixels which are spectrally homogenous and spatially contiguous are then selected. The training areas are chosen to be representative of the spectral variation over the area of interest. The training classes are then clustered, using canonical variate analysis, and representative classes are identified as reference spectral classes for a maximum likelihood classification. Posterior probabilities of class membership and typicality indices are used to identify regions which are poorly allocated or atypical, and to define further spectral classes, the steps being iterated until no further classes are needed.

CHOICE OF BANDS FOR THREE-COLOUR DISPLAY

The selection and extraction of training classes is made easier by suitable three-colour displays which enhance class differences. Probably the single most useful approach is to plot the digital counts against the band number (or perhaps against wavelength), especially if this is done systematically over the area of interest.

Such a display is also an important aid in assessing the homogeneity of training classes in those spectral regions which are not included in the three-colour display.

Techniques such as principal component analysis and decorrelation stretch are also sometimes used to enhance differences in the image.

Decorrelation stretch (Gillespie et al, 1986) provides enhanced three-colour displays when the spectral bands are highly correlated. The approach involves three stages: a principal component transformation or rotation; a scaling or stretching of these transformed values; and a reversal of the original rotation. If the second-stage scaling is such that the resulting sets of scores have equal variance, then the overall transformation can be shown to result in linear spectral indices which are uncorrelated and which have equal variances. Linear combinations of the spectral bands are produced; these are simply weighted sums and differences. Hence the indices can be interpreted directly in terms of spectral contrasts.

CANONICAL VARIATE ANALYSIS

Successful mapping of cover classes depends on the spectral data providing separation between the classes. Ideally, representative training site data will form discrete clusters corresponding to cover classes of interest.

Canonical variate analysis is a powerful ordination procedure which can be used to measure the spectral separation between sites, and to display the locations of the training site data in spectral space (see, e.g., Schowengerdt, 1982, p. 165). The procedure determines that linear

combination of the bands - that spectral index - which best separates the classes of interest, in the sense that the ratio of the between-class variation to the within-class variation is a maximum. Details can be found in Campbell and Atchley (1981). The analysis can be improved by using robust statistical procedures to downweight the effect of atypical pixels.

Successive canonical vectors are chosen to maximise the between-to-within class ratio, subject to the variate scores being uncorrelated with those already chosen. These ratios are referred to as canonical roots. Their sum measures the overall separation between the classes; the individual values indicate the relative importance of the successive vectors. These values summarise the adequacy of the data to separate the classes, and the essential dimensionality of the class separation.

It is often possible and instructive to simplify the canonical vectors without reducing the degree of class separation. The effect of reducing the set of bands can also be calculated. When this type of analysis is directed to particular class contrasts, further insight into the nature of the spectral separation can be gained.

SPECTRAL-TEMPORAL INDICES

The interpretation of the results from multi-temporal analyses can often be simplified if the canonical vectors can be constrained so that the coefficients of the spectral bands are in the same proportion within each overpass, while at the same time maintaining the degree of class separation (Campbell and Kiiveri, 1988). The interpretation then concerns the relative weightings for bands and times, resulting in a simplified spectral-temporal index. The actual procedure reduces to a between-to-within classes analysis of a linear combination of the spectral bands over time, followed by a similar analysis of a linear combination of the times over the spectral bands; the procedure is iterated until convergence.

In a study on discrimination between crop and pasture using multitemporal MSS data, a

simplified spectral-temporal index gave similar separation to that of the usual canonical variate analysis. However, the optimal index, and even the discriminating bands, were shown to vary with local cultural conditions across farm boundaries. The message from such an analysis is clear - indices derived from one analysis may be quite inappropriate in a different context.

ENHANCED CLASSIFICATION PROCEDURES

Allocation results are generally presented in the form of a map of class-labelled pixels. The maximum likelihood classifier based on multivariate Gaussian densities is often used. In this procedure, the relative (posterior) probability of membership is calculated for each known class, and the pixel is labelled as belonging to the class with the greatest value. The posterior probability for a pixel for a particular class depends on the prior probability for the class and the value of the estimated multivariate Gaussian density for the pixel for the class (Richards, 1986, p. 175). Specifically:

Posterior Probability for class k given pixel $x =$
 $(\text{Prior Probability for class } k) \times (\text{multivariate}$
 $\text{Gaussian density for pixel } x_{\text{pm}} \text{ for class } k)$
 $+ (\text{sum over all classes of the prior probability}$
 $\times \text{Gaussian density}).$

Such a summary is adequate only if all classes are well separated, and each pixel falls clearly into a known class.

In our experience, it is preferable to calculate for each pixel two sets of indices: the relative or posterior probabilities of class membership; and the typicality probabilities of belonging to the known spectral classes. The two sets of probabilities reflect different aspects of a classification. A typicality probability indicates whether a pixel is likely to belong to the corresponding training class. Each index is interpreted in a statistical sense as a tail-area probability; a pixel which is distant from a training class will have a small typicality index (Aitchison et al, 1977). The class membership probabilities are the relative values of belonging to the training classes, on the

assumption that a pixel actually belongs to one or other of the classes. The typicality index needs to be examined to check that such an assumption is reasonable.

These values may be displayed in various ways to assess the completeness and adequacy of the allocation results.

In practice, three types of display have been found to be particularly useful. Assignment of class probabilities to the red, green and blue guns allows equivocally-labelled pixels to be readily identified as mixed colours in the display. The confidence in the allocation results for each class in turn can be summarised by a display of the class probability in blue, and the typicality index in the other colours; pixels which have high probability and typicality for the class will be white in this display, while blue indicates that the pixel is closer to that class than to the other known classes, but is atypical of the training class data. The display of maximum likelihood class labels is modified by the application of a typicality threshold; pixels which are atypical of all known classes are then displayed in black. The advantage of calculating the typicality index is that various thresholds can be applied to the displays of the class labels and posterior probabilities. (The threshold could also be made class-dependent.)

Side-by-side displays of the original image data and the allocation results allow ready identification of areas which are allocated equivocally, or which are atypical. These areas may be chosen as training sites for iterative refinement of the classification.

CONTEXTUAL OR NEIGHBOUR CLASSIFICATION

There is often a considerable amount of contextual information in neighbouring pixels which may be used to improve the information on a central pixel, and hence the overall classification accuracy. One such procedure is an iterative one in which the posterior probabilities are recalculated after each classification using local prior weights. These

weights depend on the labels of the neighbouring pixels and their associations in the previous classification (see, e.g., Besag, 1986, and Discussion therein). Specifically, the calculation of the posterior probability given above is modified by replacing the (Prior Probability for class k) by a (Local Prior Probability for class k) (and similarly in the denominator). A simple model takes the logarithm of the local prior probability as a linear function of the number of pixels labelled as class k in the surrounding 3×3 or 5×5 neighbourhood of the central pixel.

Where large areas of spectrally-close classes are found, considerable overall improvement may result. The resulting classification map is also generally more appealing visually. The improvements also depend on the initial classification being reasonably accurate. Updating from a poor initial classification may actually reduce overall accuracy. In an example using MSS data for classification of crop, pasture and bush areas, incorporation of neighbour information improved overall accuracy of maximum likelihood labels from 80.7% to 94.5%.

RELATED IMAGE PROCESSING SOFTWARE

The statistical and collaborative projects carried out by the group have resulted in the development of simple-to-use image processing software for the Commodore Amiga. Modules for three-colour display of images (featuring choice of various indices for each colour gun and various filters), for split-screen display of the images, and for grey-level and pseudo-colour display of single-band images, have been developed. A related classification module consists of separate routines for discrimination between cover classes using canonical variate analysis; for band selection; and for maximum likelihood classification. The latter routine optionally provides the relative likelihoods (or posterior probabilities) of membership of each reference class, together with the typicality indices to indicate the closeness to each of the reference classes. A full-screen or split-screen module displays both original and classified

images. Options for the classification window include a display of class labels; of class labels with masking of thresholded pixels; and of the posterior probability (with thresholding) for each class in turn. A hardware-and-software package for the Amiga is now available commercially.

REFERENCES

- Aitchison, J., Habbema, J.D.F. and Kay, J.W., "A critical comparison of two methods of statistical discrimination", *Appl. Statist.*, Vol. 26, No. 1, pp. 15-25, 1977.
- Besag, J., "On the statistical analysis of dirty pictures", *J. Roy. Statist. Soc. B*, Vol. 48, No. 3, pp. 259-302, 1986.
- Campbell, N.A. and Atchley, W.R., "The geometry of canonical variate analysis", *Syst. Zool.*, Vol. 30, No. 3, pp. 268-280, 1981.
- Campbell, N.A. and Klieri, H.T., "Spectral-temporal indices for discrimination", *Appl. Statist.*, Vol. 37, No. 2, pp. 51-62, 1988.
- Gillespie, A.R., Kahle, A.B. and Walker, R.E., "Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches", *Remote Sens. Environ.*, Vol. 22, pp. 343-365, 1987.
- Richards, J.A., "Remote Sensing Digital Image Analysis", Springer-Verlag, 1986.
- Schowengerdt, R.A., "Techniques for Image Processing and Classification in Remote Sensing", Academic Press, 1983.

SPATIAL-TEMPORAL AUTOCORRELATED MODEL FOR CONTEXTUAL CLASSIFICATION OF SATELLITE IMAGERY

Nahid Khazenie and Melba M. Crawford

Center for Space Research
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

Most statistical classification techniques have assumed independence of concurrent feature vectors in remote sensing data. However, this assumption may only be valid for very large pixel sizes. As the spatial resolution of images increases, this assumption needs to be relaxed. In addition, all existing procedures ignore any temporal information in the image data. Some satellite systems cover the globe twice daily. Most processes which are observed remotely from these satellites exhibit a temporal correlation structure over a time scale which is longer than this period. This temporal correlation structure provides useful information which should result in more accurate classification of many types of processes.

A method is developed for contextual classification which considers both spatial and temporal correlation for processes which satisfy second order stationarity conditions. The feature vectors are expressed as the sum of two independent processes, one having a class dependent distribution, the other being a contaminating noise process which is autocorrelated both in space and time. The new algorithm is implemented, and results from classification of Monte Carlo simulations are summarized.

I. INTRODUCTION

One of the first attempts to incorporate contextual information involved enlarging the data set to include the observations on the four orthogonally adjacent pixels [1, 2], where the pixels located in the North, East, South, and West of the central pixel were also used to classify the central pixel. A linear combination of these five vectors was used in linear discriminant analysis to introduce some spatial smoothing prior to classification [3]. This approach yields improved results if there are large homogeneous areas where boundary perimeters are small relative to the area of the region.

The problem with these procedures is that neighboring pixels may well belong to different classes. This can be handled properly only if a model for the joint distribution of all class labels is known. Unfortunately, this results in an impractical optimization problem. The conditional joint probability distribution can be simplified by allowing for contextual information only through the four horizontal and vertical neighbors, [4, 5, 6]. However, totally general boundary configurations cannot be specified under this assumption. Owen, Hjort, and Mohn assume at most two labels A and B in a neighborhood and allow only configurations of the form where

at most one class boundary passes through a neighborhood. The assumption reduces the number of possible patterns considerably. If some pixels have known labels (training data), the required parameters can be estimated from the data.

Hjort, Mohn, and Storvik [9] generalize this formulation by assuming a Gaussian model where the process governing the vector of intensities is the sum of two independent processes, one having a class dependent structure, and the other being a contaminating noise process containing the correlation structure. They are able to account for spatial correlation between neighboring sites with this more realistic model. Khazenie [8] extends this approach to account for both spatial and temporal correlation structures. The procedure is outlined, and results of simulation experiments are presented in the following sections.

II. STATISTICAL CLASSIFICATION APPROACH

The image being classified consists of N pixels, where X_i is the observed vector of intensities for pixel i , $i = 1, \dots, N$. The components of this d -dimensional vector are typically spectral measurements in different channels as well as other topographic information. The image consists of K classes, each having prior probability $\pi(k)$, $k = 1, \dots, K$. The X vectors from class k are distributed according to a density $f_k(x)$, which is assumed to be Gaussian.

A statistical classification procedure estimates the true class label L_i of pixel i based on the observed X_i 's. Some preliminary knowledge about the class distributions is required. This prior knowledge, typically in the form of statistical parameters, is estimated using a training set for each class. In general, classification procedures then produce assignments of class labels L_i 's to pixels. It is also possible that the assigned class label is the doubt class D . Therefore, $L_i \in \{1, \dots, K, D\}$.

Let the cost function associated with pixel i when L_i is assigned have the form

$$C(L_i^*, L_i) = \begin{cases} 0 & \text{if } L_i^* = L_i \\ 1 & \text{if } L_i^* \neq L_i \text{ \& } \neq D \\ c & \text{if } L_i^* = D \end{cases} \quad (1)$$

where often $c \in [0, (1 - \frac{1}{K})]$.

It can be shown that an approximation to Bayes rule is

$$L_i = \begin{cases} D & \text{if } P_i(k | \Delta_i) \leq 1 - \epsilon \quad \forall k = 1, \dots, K \\ m & \text{if } P_i(m | \Delta_i) = \max_{k \in (1, K)} P_i(k | \Delta_i) \end{cases} \quad (2)$$

where

$$P_i(k | \Delta_i) = \Pr \{L_i = k | \Delta_i\} \quad (3)$$

and Δ_i is a suitably chosen set of observations including at least $\{X_i\}$. The natural choice of most contextual classifiers is

$$\Delta_i = \{X_{iN}, X_{iE}, X_{iS}, X_{iW}, X_i\}, \quad (4)$$

so the observations from the four immediate orthogonal neighbors, North, East, South, and West are also used to classify pixel i

The posterior probabilities are

$$P_i(k | \Delta_i) = (1/f(\Delta_i)) \pi(k) \sum_I g(I | k) f(\Delta_i | k, I) \quad (5)$$

where $I = \{a, b, c, d\}$ is the neighborhood class configuration respectively for $\{iN, iE, iS, iW\}$, and $\pi(k)$ is the prior probability for class k . I is one of K^4 possible configurations, and $g(I | k)$ is the probability of seeing that configuration given the class label of central pixel i is k , (i.e., $L_i = k$). Further, $f(\Delta_i | k, I)$ is the conditional joint density of five vectors of pixel values, given that k, a, b, c, d are the classes of the neighborhood pixels. The conditional distribution of X_i , given the classes in the neighborhood, is assumed to only depend on the value and class of the central pixel i . So,

$$f(\Delta_i | k, I) = f_k(X_i) h(X_{iN}, \dots, X_{iW} | X_i, k, I) \quad (6)$$

by substitution,

$$P_i(k | \Delta_i) = A \pi(k) f_k(X_i) R_k(\Delta_i) \quad (7)$$

where A is a normalizing constant and

$$R_k(\Delta_i) = \sum_I g(I | k) h(X_{iN}, \dots, X_{iW} | X_i, k, I). \quad (8)$$

$R_k(\Delta_i)$ can be viewed as an adjustment factor which accounts for spatial dependence. This model has motivated a variety of contextual classification methods. They differ by how g , f , and h are modeled. Most researchers assume conditional independence of neighboring spectral vectors given the classes of the central pixel and the neighborhood. This assumption is inadequate when applied to high resolution data [9]. Remote sensing data from satellites are being collected at increasingly higher spatial and temporal resolution. Thus, it is important to develop classification algorithms which explicitly account for this correlation. A model for the spectral vectors that allows only positive correlation between spectral measurements from neighboring pixels, given the classes, was developed which combines both types of contextual information: the choice of g and spectral spatial-temporal correlation.

Spatial-Temporal Autocorrelated Model for Spectral Vectors

Assume the observed process is the sum of two independent processes: i) Y_i^t , $i = 1, \dots, M$ having a class-dependent structure, and ii) the autocorrelated noise process

ϵ_i^t , $i = 1, \dots, M$. The Y_i^t 's contain information about the pixel's class, and ϵ_i^t results from a variety of sources of error which give rise to the spatial and temporal dependency structure. The time index t varies from 1 to T . Let

$$X_i^t = Y_i^t + \epsilon_i^t \quad i = 1, \dots, M \quad (9)$$

where (Y_1^t, \dots, Y_M^t) and $(\epsilon_1^t, \dots, \epsilon_M^t)$ are independent, and $Y_i^t \sim N(\mu_{Li}, (1 - \theta)\Sigma)$ and $(\epsilon_1^t, \dots, \epsilon_M^t)$ are multinormally distributed with zero mean and covariance function,

$$\text{COV}(\epsilon_i^t, \epsilon_j^{t(v)}) = \rho_s^{|i-j|} \rho_v^v \theta \Sigma \quad (10)$$

such that ρ_s is the spatial correlation parameter, and $|i - j|$ is the ordinary euclidean distance between centers of pixels in an integer-valued coordinate system. The value of θ determines the fraction of the variance of X_i^t attributed to ϵ_i^t . Assuming an exponential decay in the correlation process, the lag one temporal correlation is ρ_v , and v is the time lag between time t and $t^{(v)}$. Therefore, $X_i^t \sim N(\mu_{Li}, \Sigma)$ given L_i . The X_i^t 's are correlated, accounting for both spatial correlation ρ_s and temporal correlation ρ_v :

$$\begin{aligned} \text{COV}(X_i^t, X_j^{t(v)}) &= \rho_s^{|i-j|} \rho_v^v \Sigma \quad \text{for } i \neq j, v \neq 0, \\ &\quad \rho_s \neq 0, \rho_v \neq 0 \\ &= \rho_v^v \Sigma \quad \text{for } \rho_s = 0 \\ &= \rho_s^{|i-j|} \Sigma \quad \text{for } \rho_v = 0 \end{aligned} \quad (11)$$

This formulation allows for temporal and spatial dependence to exist simultaneously or individually, and the value of θ defines the level of contamination of the correlated noise.

To construct the desired classification rule involving $f(\Delta_i | k, I)$ in (5), it is necessary for Δ_i to include temporal information. Let Δ_{it} denote pixel i and its four orthogonal neighbors at time t , then

$$\Delta_i = (\Delta_{it} \cup \Delta_{i(t-1)} \cup \dots \cup \Delta_{i(t-n)}) \quad (12)$$

where $t^{(n)} = t - n$, and n is the highest temporal lag where the correlation is nonzero. For notational simplicity, define

$$X_{iN}^* = \begin{bmatrix} X_{iN}^t \\ X_{iN}^{t(1)} \\ \vdots \\ X_{iN}^{t(n)} \end{bmatrix}, \dots, X_i^* = \begin{bmatrix} X_i^t \\ X_i^{t(1)} \\ \vdots \\ X_i^{t(n)} \end{bmatrix} \quad (13)$$

with $(n+1)d$ dimensional mean vectors μ_{iN}^* , μ_i^* , \dots , μ_M^* and covariance Σ^* , respectively. For a first order autoregressive

process, $f(\Delta_1 | k, I)$ is distributed multinormal in $(n+1)5d$ dimensions:

$$N_{(n+1)5d} \left\{ \begin{bmatrix} \mu_a^* \\ \mu_b^* \\ \mu_c^* \\ \mu_d^* \\ \mu_k^* \end{bmatrix}, \Sigma_{\Delta_1} \otimes \begin{bmatrix} 1 & \rho_1 & \rho_1^2 & \dots & \rho_1^n \\ & 1 & \rho_1 & \dots & \rho_1^{n-1} \\ & & \ddots & & \vdots \\ & & & \ddots & \rho_1 \\ & & & & 1 \end{bmatrix} \right\} \quad (14)$$

where Σ_{Δ_1} denotes the spatial correlation structure at lag zero and $\otimes \triangleq$ Kronecker product. For a neighborhood system where neighborhood classes are a, b, c, d, and k,

$$\Sigma_{\Delta_1} = \begin{bmatrix} \Sigma & \beta\Sigma & \gamma\Sigma & \beta\Sigma & \alpha\Sigma \\ & \Sigma & \beta\Sigma & \gamma\Sigma & \alpha\Sigma \\ & & \Sigma & \beta\Sigma & \alpha\Sigma \\ & & & \Sigma & \alpha\Sigma \\ & & & & \Sigma \end{bmatrix} \quad (15)$$

α = correlation of first order neighbors = $\rho_1 \theta$

β = correlation of diagonal neighbors = $\rho_1^{\sqrt{2}} \theta$

γ = correlation of second order neighbors = $\rho_1^2 \theta$

If $n = 1$, i.e., only lag one temporal correlation is significant,

$$\Sigma_{\Delta_1} = \Sigma_{\Delta_1} \otimes \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \quad (16)$$

The joint conditional density of $(X_{iN}^*, X_{iE}^*, X_{iS}^*, X_{iW}^*)$ given X_i^* and the classes, in (6) - (8), is also required.

$$h(X_{iN}^*, \dots, X_{iW}^* | X_i^*, k, I) \sim N_{(n+1)4d}(\mu_k, \Sigma_k) \quad (17)$$

Expressions for the mean and covariance structure of the joint conditional distribution, μ_k and Σ_k can be derived directly and are contained in [8]. For this model the posterior probability of the class given the neighborhood values is then

$$P_i(k | X_i^*, X_{iN}^*, \dots, X_{iW}^*) = A \pi(k) f_k(X_i^*) \exp\{-2\alpha^2 t^* (X_i^* - \mu_k^*)' \Sigma_k^{-1} (X_i^* - \mu_k^*)\} \sum_I g(I | k) R_i(I) \exp\{\alpha t^* (X_{iN}^* + \dots + X_{iW}^* - \mu_a^* - \dots - \mu_d^*)' \cdot \Sigma_k^{-1} (X_i^* - \mu_k^*)\} \quad (18)$$

where

$$R_i(I) = \exp\{-1/2 Z_i^{*'} \Sigma_k^{-1} Z_i^*\} \quad (19)$$

and

$$Z_i^* = \begin{bmatrix} X_{iN}^* - \mu_a^* \\ X_{iE}^* - \mu_b^* \\ X_{iS}^* - \mu_c^* \\ X_{iW}^* - \mu_d^* \end{bmatrix} \quad (20)$$

The normalizing constant A can be evaluated such that the sum of probabilities is one.

Any probability distributions chosen to model $g(I | k)$ in (18) will result in a classification procedure of the form (2). The general formula would involve K^4 possible configurations, so only choices of $g(I | k)$ which can be implemented in practice should be considered.

Modeling of Configuration Probability Distribution

The probability of occurrence of many configurations $g(a, b, c, d | k)$ is so small that they can be set to zero. A geometric probability model was proposed by Owen and Switzer [11] and then generalized by Hjort [12]. This model allows positive probability for three configurations. These three configurations are referred to as X, T and L patterns.

| | | |
|-------|-------|-------|
| k | k | k |
| k k k | k k m | k k m |
| k | k | m |

This is a realistic assumption when the regions are large compared to the pixel sizes, so that most first order neighbors are formed from only two classes. There are four possible L and four possible T patterns. Let

$$P(X) = p \quad P(L) = q \quad P(T) = r \quad (21)$$

then $p + q + r = 1$.

For an X pattern, L pattern, and T pattern respectively,

$$\begin{aligned} g(k, k, k, k | k) &= p + (q + r) \pi(k) \\ g(k, k, m, m | k) &= g(m, k, k, m | k) = g(m, m, k, k | k) \\ &= g(k, m, m, k | k) = 1/4 q \pi(m) \\ g(k, k, m, k | k) &= g(k, k, k, m | k) = g(m, k, k, k | k) \\ &= g(k, m, k, k | k) = 1/4 r \pi(m) \end{aligned} \quad (22)$$

where m is one of the remaining $K - 1$ classes and $\pi(k)$ are again the prior probabilities of the class labels. This reduces the total number of terms from K^4 to $8K - 7$, so computation time depends linearly on K. The posterior distribution of K in (18) is thus

$$\begin{aligned} P(k | X_i^*, X_{iN}^*, \dots, X_{iW}^*) &= A \pi(k) f_k(X_i^*) \exp\{-2\alpha^2 t^* (X_i^* - \mu_k^*)' \Sigma_k^{-1} (X_i^* - \mu_k^*)\} \cdot \\ &[p \exp\{4\alpha t^* (\bar{X}_i^* - \mu_k^*)' \Sigma_k^{-1} (X_i^* - \mu_k^*)\} R_i(k, k, k, k) \\ &+ q \exp\{2\alpha t^* (\bar{X}_i^* - \mu_k^*)' \Sigma_k^{-1} (X_i^* - \mu_k^*)\} G_{i,k}(X_i^*, \bar{X}_i^*) \\ &+ r \exp\{3\alpha t^* (\bar{X}_i^* - \mu_k^*)' \Sigma_k^{-1} (X_i^* - \mu_k^*)\} H_{i,k}(X_i^*, \bar{X}_i^*)] \end{aligned} \quad (23)$$

where

$$G_{1k}(X_1^*, \bar{X}_1^*) = \sum_{m=1}^K \pi(m) \exp\{2\alpha t^* (\bar{X}_1^* - \mu_k^*)' \Sigma_k^{-1} (X_1^* - \mu_k^*)\} \cdot \frac{1}{4} [R_1(k, k, m, m) + R_1(m, k, k, m) + R_1(m, m, k, k) + R_1(k, m, m, k)]$$

$$H_{1k}(X_1^*, \bar{X}_1^*) = \sum_{m=1}^K \pi(m) \exp\{\alpha t^* (\bar{X}_1^* - \mu_k^*)' \Sigma_k^{-1} (X_1^* - \mu_k^*)\} \cdot \frac{1}{4} [R_1(k, k, k, m) + R_1(k, k, m, k) + R_1(k, m, k, k) + R_1(m, k, k, k)]$$

and $\bar{X}_1^* = \frac{1}{4}(X_{1N}^* + X_{1E}^* + X_{1S}^* + X_{1W}^*)$

For some processes the covariance Σ given by the initial modeling of $X_1 \sim N(\mu_{L_1}, \Sigma)$ used throughout the development of the methodology is class dependent, i.e. $X_1 \sim N(\mu_{L_1}, \Sigma_{L_1})$. This extension is straightforward. The procedure can be considered as a best linear rule leading to contextual and temporal improvements and as a best quadratic rule corresponding to different prior covariance matrices $\Sigma_1, \dots, \Sigma_K$.

Unbiased minimum variance estimates were calculated for all indicated parameters. Data for Nonoverlapping neighboring sites were used as training sites from the exterior of clearly homogeneous areas to estimate all spatial parameters in order to assure the independence of the sites.

III. SIMULATION RESULTS

The algorithm was implemented in Fortran and run on a Dual Cyber 170/750 computer.

First order space-time processes were simulated where all parameters and ground truth were known. To investigate the effect of several parameter sets, some extreme cases were considered. Case I had considerable first order spatial and temporal correlation. Case II was created with high temporal correlation, but very low spatial correlation. Case III was generated to have low temporal correlation and high spatial correlation. The scenes in the three simulations have dimension 40×40 , a two-channel spectral vector, and three classes. A Landsat scene with three different geological classes was used as a ground truth. One hundred and eighty scenes were generated for each simulation. Tables 1, 2, and 3 contain lists of the design and the corresponding estimated parameter sets for the first, second, and third cases respectively.

For each of the three simulations, seven of the one hundred eighty scenes were selected systematically for classification. The classification results of the seven scenes using the spatial temporal autocorrelated model are summarized in the remainder of this section.

Figures 1 - 3 show the percentage of misclassified pixels which resulted from estimating the temporal correlation and utilizing it in the classification, compared to using only spatial information for different parameter values. There is a definite improvement which clearly depends on the amount of temporal correlation in the process. It is interesting that the relation between percentage improvement and the magnitude of ϕ is approximately exponential. This is better than the expected linear improvement.

The performance of classifier was evaluated for different values of both spatial and temporal correlation in the pro-

cess. Figure 4 illustrates these results for the three first-order processes. This figure shows that spatial-temporal statistical classifier performed much better when spatial correlation was low and temporal correlation was high. Results were not as good when both types of correlation were large. It appears that the temporal correlation structure is modeled better than spatial correlation. This could be explained by the fact that temporal correlation is a contaminating source in only one direction in time, and it is modeled to affect only the same site. Spatial correlation structure is more complex. Spatial correlation contaminates in two directions and affects the site and all of its first order neighbors.

Overall average error rates can be summarized as:

$$\begin{aligned} \rho_s = 0.40, \phi = 0.65 &\Rightarrow 20.9 \% \text{ error} \\ \rho_s = 0.05, \phi = 0.95 &\Rightarrow 17.5 \% \text{ error} \\ \rho_s = 0.40, \phi = 0.20 &\Rightarrow 18.3 \% \text{ error} \end{aligned}$$

IV. CONCLUSIONS

Illustrations, in the form of images, are useful in evaluating some additional performance measures of the new classification procedure. Figure 5 shows the classified images of statistical classifier for $\phi = 0.65$ and $\rho = 0.45$. The statistical classifier performs quite well near borders between classes as well as inside homogenous regions. Other classifiers which were investigated did not depict the shape of homogeneous regions as well.

The contextual classification procedure, which was developed, models both spatial and temporal correlation structures for stationary processes. The mathematical structure of the new methodology allows for diverse modeling of the joint class distributions. The temporal correlation structure is also quite general. Although the examples were generated only for the case where first order temporal correlation existed, the formulation of (18) and (23) is not limited to this form. The mathematical formulation supports the general case of $\rho_v^{(v)} = \phi_v$, which is still computationally feasible, although more complicated to implement than the first order process used in the examples. More general models of the correlation structure and the probability model of the class distribution are currently being investigated.

BIBLIOGRAPHY

- [1] P. Switzer, "Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery," *Math. Geol.*, vol. 12, pp. 367-376, 1980.
- [2] P. Switzer, "Some spatial statistics for the interpretation of satellite data," *Bull. Int. Stat. Inst.*, vol. 50, no. 2, pp. 962-971, 1983.
- [3] P. Switzer, W. S. Kowalik, and R. J. P. Lyon, "A prior probability method for smoothing discriminant classification maps," *Math. Geol.*, vol. 14, pp. 433-444, 1982.
- [4] A. Owen, "A neighborhood-based classifier for Landsat data," *Canadian J. Statist.*, vol. 12, pp. 191-209, 1984.
- [5] N. L. Hjort, and E. Mohn, "A comparison of some contextual methods in remote sensing classification," *Proc. 18th Int. Sym. Remote Sensing Env.*, CNES, Paris, 1984.
- [6] J. Haslett, "Maximum likelihood discriminant analysis in the plane using a Markovian model of spatial context," *Patt. Recog.*, 1985.

- [7] N. L. Hjort, "Neighborhood based classification of remotely sensed data based on geometric probability models," Dept. of Statistics, Stanford University, Technical Report 10, 1985.
- [8] N. Khazenie, and M. M. Crawford, "Contextual Classification of Remotely Sensed Data, Based on a Spatial-Temporal Correlation Model," *Proceedings of the AVHRR Users Meeting*, Boulder, Colorado, July 1987.
- [9] N. L. Hjort, E. Mohn, and G. Storvik, "Contextual classification of remotely sensed data based on an autocorrelation model," *Norwegian Computing Center*, Technical Report No. 13, 1985.
- [10] T. W. Anderson, *Introduction to Multivariate Analysis*, 2nd ed. New York: Wiley, 1984.
- [11] A. Owen, and P. Switzer, "A neighborhood-based classifier for Landsat data," *Dept. of Stat., Stanford University*, Technical Report, 1982.
- [12] N.L. Hjort, "Estimating parameters in neighborhood based classifiers for remotely sensed data, using unclassified vectors," *Norwegian Computing Center*, Technical Report No. 12, 1985.

TABLE 1 SIMULATION ONE
Substantial Spatial and Temporal Correlation

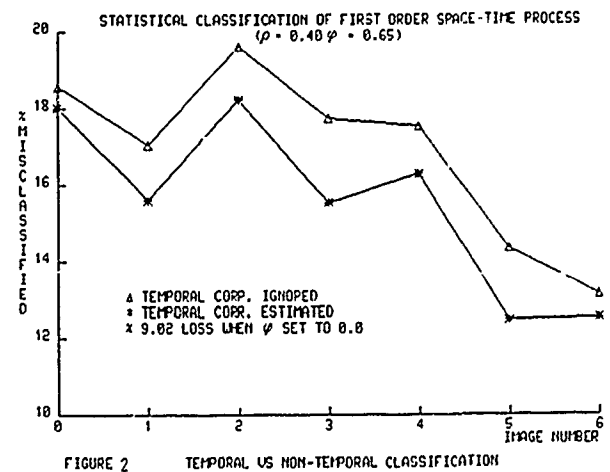
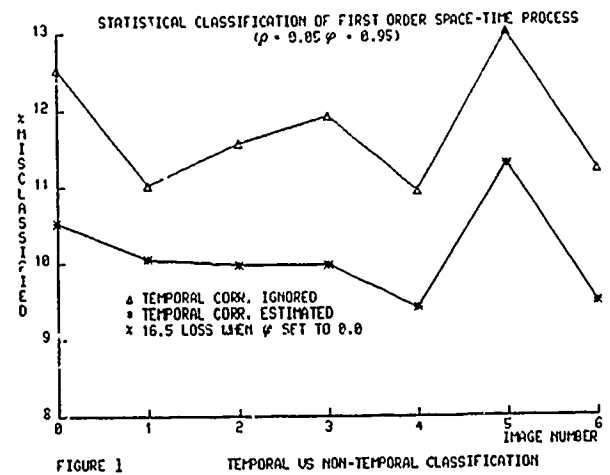
| PARAMETERS | DESIGN VALUES | ESTIMATED VALUES |
|----------------|--|--|
| $\mu(d, k)$ | $\begin{pmatrix} 0 & 2 & 4 \\ 0 & 3 & 6 \end{pmatrix}$ | $\begin{pmatrix} 0.01629 & 1.981 & 3.98 \\ 0.0096 & 3.02 & 6.01 \end{pmatrix}$ |
| $\sum_k(d, d)$ | $\begin{pmatrix} 2.0 & 1.5 \\ 1.5 & 4.0 \end{pmatrix}$ | $\begin{pmatrix} 1.99128 & 1.50137 \\ 1.50137 & 3.9655 \end{pmatrix}$ |
| ρ_s | 0.400 | 0.38860 |
| ϕ | 0.650 | 0.64338 |
| θ | 0.850 | 0.84367 |
| α | 0.340 | 0.32490 |
| β | 0.233 | 0.22170 |
| γ | 0.136 | 0.12180 |

TABLE 2 SIMULATION TWO
Low Spatial and High Temporal Correlation

| PARAMETERS | DESIGN VALUES | ESTIMATED VALUES |
|----------------|--|--|
| $\mu(d, k)$ | $\begin{pmatrix} 0 & 2 & 4 \\ 0 & 3 & 6 \end{pmatrix}$ | $\begin{pmatrix} 0.016 & 1.995 & 3.93 \\ 0.079 & 3.03 & 5.911 \end{pmatrix}$ |
| $\sum_k(d, d)$ | $\begin{pmatrix} 2.0 & 1.5 \\ 1.5 & 4.0 \end{pmatrix}$ | $\begin{pmatrix} 1.915 & 1.576 \\ 1.576 & 4.055 \end{pmatrix}$ |
| ρ_s | 0.05000 | 0.10070 |
| ϕ | 0.95000 | 0.93700 |
| θ | 0.05000 | 0.00191 |
| α | 0.00250 | 0.00590 |
| β | 0.00070 | 0.00945 |
| γ | 0.00012 | 0.00335 |

TABLE 3 SIMULATION THREE
High Spatial and Low Temporal Correlation

| PARAMETERS | DESIGN VALUES | ESTIMATED VALUES |
|----------------|--|---|
| $\mu(d, k)$ | $\begin{pmatrix} 0 & 2 & 4 \\ 0 & 3 & 6 \end{pmatrix}$ | $\begin{pmatrix} -0.004 & 2.016 & 3.993 \\ 0.015 & 3.003 & 5.979 \end{pmatrix}$ |
| $\sum_k(d, d)$ | $\begin{pmatrix} 2.0 & 1.5 \\ 1.5 & 4.0 \end{pmatrix}$ | $\begin{pmatrix} 1.989 & 1.487 \\ 1.487 & 3.979 \end{pmatrix}$ |
| ρ_s | 0.4000 | 0.3909 |
| ϕ | 0.2000 | 0.1960 |
| θ | 0.9000 | 0.8960 |
| α | 0.3600 | 0.3503 |
| β | 0.2463 | 0.2374 |
| γ | 0.1440 | 0.1338 |



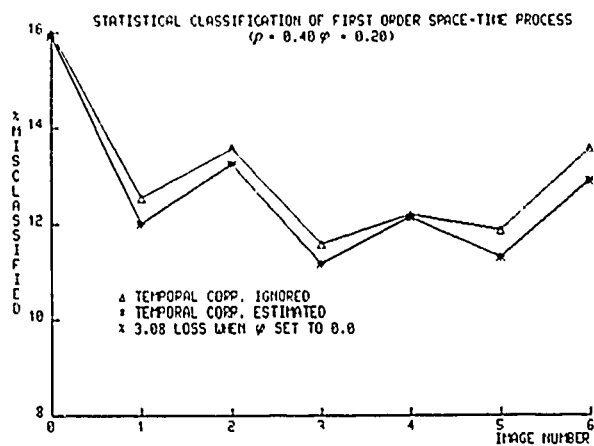


FIGURE 3 TEMPORAL VS NON-TEMPORAL CLASSIFICATION

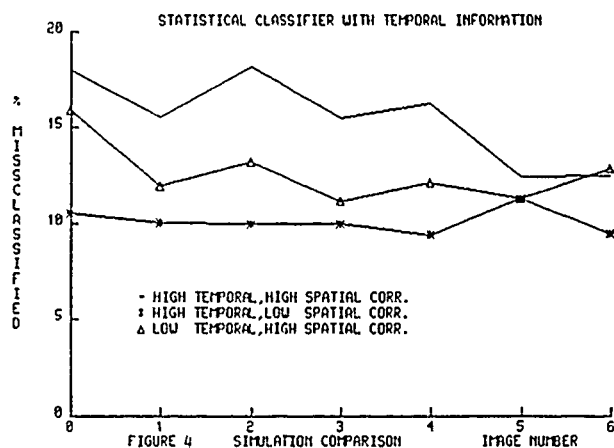
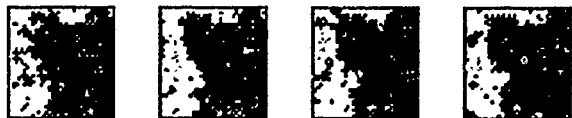


FIGURE 4 SIMULATION COMPARISON

SIMULATED PROCESSES WITH HIGH TEMPORAL & HIGH SPATIAL CORR.



INTENSITY VALUES FOR CHANNEL ONE FOR 4 IMAGES



CLASSIFIED IMAGES



GROUND TRUTH

TEMPORAL CORR. LAG-1, $\varphi = 0.65$
 SPATIAL CORR. LAG-1, $\rho = 0.40$

FIGURE 5 STATISTICAL CLASSIFICATION, TEMPORAL CORR. ESTIMATED

EIGENVECTOR PROJECTION TRANSFORMATION AND DIMENSION SIZE REDUCTION IN REMOTE SENSING DATA PROCESSING

Bo Li and Wooil M. Moon

Geophysics, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

ABSTRACT

One problem in multidimensional remote sensing data processing is the reduction of information space from m -D to 3-D for RGB colour display and visual analysis with minimal loss of important information. Principal Component Analysis (PCA) has been used, but the resulting 3 components are not correlated and can be considerably different in terms of information significance. A new projection transformation was tested in this paper. In this approach, information structure of the data set is analyzed by using eigen analysis, followed by Householder transformation to establish a transformation matrix. Finally the original m -D data set is transformed into a new lower dimensional feature space, in which each feature has the same degree of significance. To enhance the visual display of the resultant data, techniques for scatter adjustment and rotation projection are also applied and tested. Test of the method with Landsat TM data on Sudbury Area, Ontario was carried out for geological applications. The results indicate this method is effective and feasible for routine application.

KEYWORD: eigen analysis, orthogonal projection, image transformation, data processing, reduction of dimension.

1. INTRODUCTION

With increasing number of spectral bands being used in satellite and air-borne remote sensing data and the development of integrated data processing, there is more concern for the problem of how to reduce dimension size of remote sensing data sets while maintaining the visual information [1][2][3]. To achieve this, emphasis has been placed on either the statistical feature selection, or on the effective utilization of colour spaces [2][3]. The connection between these two is rarely investigated. For example, in spite of PCA in dimension compression of data set, the image often becomes confusing and degraded [1] when the result of PCA is displayed in RGB space. This is caused from the fact that there is no correlation among principal components computed using PCA. In terms of variance contributions which can be understood as a measure of information contained, the PCA transformed components are very different from one another. They occupy colour channels of equal weight in RGB colour space and can overlap others. In the case of Landsat TM data, certain components computed from PCA contain a high level of random and

strip noise. When such a component is incorporated into a colour composite, it will unavoidably degrade the whole image. Also, in the interpreter's point of view, correlation between different bands displayed in RGB space is very useful for accurate interpretation, although it may sometimes be desirable to decompose the highly correlated data set into non-correlated ones for individual band interpretation.

To resolve these problems, a new projection transformation method is developed and tested for dimension size reduction. By using this method, data sets can be transformed from m -D to RGB colour space while each new pseudo band has equal information content. Furthermore, since this transformation is not unique, an orthogonal rotation can be successfully applied to rotate the projection axes. This approach makes it possible for interpreters to view a colour image from different perspectives in feature space. To increase visual discrimination in feature space, scatter adjustment is also tested. For many applications in remote sensing data processing, classification is one of the important objectives. In this research, the transformed data sets using different parameters are also examined to estimate the effects of the various transformation on the classification capability.

1.1 Multidimensional Data Structure

Here "structure" refers to the distribution attribute of a data set in feature space. In most cases, it is acceptable to assume that the data set concerned observes multidimensional normal distribution. The distribution shape of a data set can be understood as a hyperellipsoid in m -D feature space. The length of each axis of the hyperellipsoid can be accordingly understood as the measurement of data deviation along this axis. In the case of remote sensing data sets, especially multispectral data sets, all spectral bands are highly correlated in the set of spectral bands. In feature space this represents the fact that the ratio of the length of the longest axis to the length of other axes is accordingly high. Subsequently, most data points will cluster around the longest axis. In statistical sense, the projection onto the longest axis is best representative of the complete data set if only one feature can be left after a dimension reduction.

1.2 Projection Transformation and Dimension Reduction

Transforming a data set from m -D feature space into 3-D color space requires that the following conditions be satisfied:

(a) After the dimension reduction, most of the origi-

- nal information must be retained.
- (b) Each of the new features has the equal information.
 - (c) Each of the new projection axes corresponding to each new feature is orthogonal to other axes.
 - (d) The amount of calculation involved in the transformation is acceptable for the given computer system.

A linear transformation approach which satisfies the above conditions has been chosen for the data space dimension reduction. The linear transformation can also be interpreted as a kind of projection transformation in m-D feature space. In actual numerical processing, the transformation is presented by a transform matrix T where, each column of T is defined as a projection axis. The inner product between the data vectors and projection axis defines the projection of the data vector onto the projection axes. Numerical implementation of this transformation can be divided into the following steps:

(i) Choice of an axis from the m-D data hyperellipsoid as the central axis around which the projection axes will establish (Fig. 1A).

(ii) Establishment of three projection axes which are orthogonal to one another, such that the angle between each of these projection axes and the central axis is same (Fig. 1A).

(iii) Projection of the data set onto the three projection axes.

In this way, three new feature bands with equally distributed information can be obtained. It is easy to understand that the attribute of a projection largely depends on the direction of the central axis in feature space. If a different axis is chosen as the central axis, the distribution and interpretable level of information contained in transformed output will also be different.

1.3 Further Rotation

It should be pointed out that the transformation which satisfies conditions in (1.2) is not unique. In fact, any established 3-axes set can be rotated around a central axis to form a new set which corresponds to a new projection angle or view angle (Fig. 1B). By changing rotation angle, the possible number of 3-axes sets is theoretically infinite. This mechanism provides interpreters with improved flexibility to examine the data set from different view angle in feature space. To keep the data set structure tractable, the rotation is limited to an orthogonal rotation.

1.4 Data Set Scatter Adjustment

When correlation among the new features is very high, which indicates most data points cluster closely to the central axis, the saturation of resultant RGB composite may decrease. To enlarge variety in RGB composite, scatter of data points can be increased for enhancement of RGB display. To achieve this, the distance between every individual data point and the central axis is firstly measured by using Euclidian distance or an angle (Fig. 2). Then, according to given adjustment factor, the distance or angle can be increased. Such an incremental move will result in pushing data points away from the central axis, and lower correlation among the new feature bands can be achieved.

1.5 Classification Capability

As it can be seen in Fig. 1B, when rotation angle θ of the projection axes around the central axis changes, the projection onto other coordinate planes will also keep changing. If there are several identifiable classes contained in the data set, the projec-

tion of each class onto coordinate plane also continuously changes. When a specific classification is considered in a chosen plane, the rotation angle of the projection axes has a direct effect on the classification result. To estimate this effect, the Fisher ratio [4,5] is used. It is defined as the ratio of the between-class scatter to the sum of the within-class scatter. It is obvious that the larger the Fisher ratio is, the better the classification results are. The optimal rotation angle should make the Fisher ratio to reach its maximum.

2. METHODOLOGY

2.1 Eigen Analysis for Data Structure Determination

Let X denote raw data matrix, $X = (x_{ij})_{n \times m}$, where i denotes the i th sample or observation and j denotes the j th feature or band, m is the number of features and n is the number of samples. Then let $\bar{X} = (\bar{x}_{ij})_{n \times m}$, where $\bar{x}_{ij} = \bar{x}_{ji}$, ($1 \leq i \leq n$) is the mean of the j th feature. The variance-covariance matrix for X can then be written as:

$$\Sigma = (X - \bar{X})^T (X - \bar{X}). \quad (2.1.1)$$

According to multivariate statistical theory, there exists an orthogonal matrix V such that:

$$V^T \Sigma V = \begin{bmatrix} \lambda_1 & 0.0 & \dots & 0.0 \\ 0.0 & \lambda_2 & \dots & 0.0 \\ \vdots & \vdots & \ddots & \vdots \\ 0.0 & 0.0 & \dots & \lambda_m \end{bmatrix} \quad (2.1.2)$$

where $\lambda_i > \lambda_{i+1}$, ($i = 1, 2, \dots, m$) are eigenvalues of and V_i , the i th column in the matrix V which corresponds to λ_i , is the i th eigenvector of Σ . Each eigenvector indicates the direction of one axis of the hyperellipsoid of data X in m-D feature space and the magnitude of the corresponding λ represents the length of the axis [1]. Eigen analysis provide a profile of the complete data set in statistical context.

2.2 Establishing Transformation Matrix using Household Method.

Based on the data set structure and conditions listed in (1.2), a linear transformation can be defined. In actual data processing, this transformation is represented by a matrix $T = (t_{ij})_{n \times r}$, which can either be called transformation matrix or projection matrix. The transformation will be accomplished by matrix multiplication:

$$Y = XT \quad (2.2.1)$$

where, $Y = (y_{ij})_{n \times r}$, ($r < m$) is the output result of the transformation and dimension reduction, r is the number of new features after transformation. When the result is going to be displayed with a RGB colour display, it is usually set $r=3$. According to (1.2), T should satisfy the following:

$$T_i \cdot T_j = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases}$$

$$T = (T_1, T_2, \dots, T_r) \quad (2.2.2)$$

where T_i , ($i = 1, 2, \dots, r$) are projection axes. For a given data structure, it is easy to understand that along the direction of the first eigenvector, the deviation of the projection of the complete data set can reach its maximum λ_1 . It is not unusual in the case of remote sensing data processing that the variance contribution in the direction of first the eig-

envektor $\lambda_1 / \sum_{i=1}^m \lambda_i$ is equal to or greater than 90%. This indicates that if T_i , ($i = 1, 2, \dots, m$) are placed around the first eigenvector V_1 or namely V_1 is chosen as the central axis, most of the original data information will be retained after the transformation $Y = XT$. To distribute the information equally, T_i should satisfy the following:

$$T_i \cdot V_1 = T_j \cdot V_1, \quad i \neq j, \quad 1 \leq i, j \leq r. \quad (2.2.3)$$

Eq. (2.2.3) represents that the angle α between any new projection axis T_i and V_1 is the same (Fig. 1A), and each of three axes has an equal information contribution through the projection transformation in feature space. Since V_1 is obtained through eigen analysis, only T_i has to be determined. Let us consider an average vector V_a which is defined as follows:

$$V_a = \frac{\sum_{i=1}^r V_i}{\|\sum_{i=1}^r V_i\|} \quad (2.2.4)$$

Notice that V_i , ($1 \leq i \leq r$) is orthogonal to each other, then

$$V_i \cdot V_j = \begin{cases} 1, & \text{if } i=j; \\ 0, & \text{otherwise} \end{cases}$$

$$V_a \cdot V_i = \frac{1}{\|\sum_{i=1}^r V_i\|} \quad (2.2.5)$$

comparing (2.2.5) with (2.2.2) and (2.2.3), it becomes obvious that if the complete data set can be rotated so that the new maximum deviation projection axis V_1^* overlaps V_a , then the V_i , ($1 < i \leq r$) can be chosen to form transformation matrix T . Instead of rotating the complete data set, we can find an orthogonal matrix H to rotate V to V^* which overlaps V_1 . Under this orthogonal rotation, V_i , ($1 < i \leq r$) will be changed to V_i^* , ($1 < i \leq r$). Because H is an orthogonal matrix, V_i^* , ($1 \leq i \leq r$) and V_a^* will satisfy the following:

$$V_a^* = V_1$$

$$V_i^* \cdot V_j^* = \begin{cases} 1, & \text{if } i=j; \\ 0, & \text{otherwise} \end{cases}$$

$$V_a^* \cdot V_i^* = V_a^* \cdot V_j^*, \quad i \neq j, \quad 1 \leq i, j \leq r \quad (2.2.6)$$

Then, T can be formed by using V_i^* , ($1 < i \leq r$), i.e. $T = (V_1^*, V_2^*, \dots, V_r^*)$. At this point, the next question is how to find the matrix H which satisfies the condition

$$H V_a = V_a^* = V_1. \quad (2.2.7)$$

In this research, Household method [6] is applied for this purpose. In the theory of linear algebra, for any given vector Z and an unit vector G in m -D vector space, there always exists a matrix H such that

$$HZ = \|Z\| \cdot G \quad (2.2.8)$$

where H is called Household matrix. The linear transformation defined by H is also called the Household transformation [6]. H is usually derived in the following way:

$$H = I - 2 \cdot UU^T$$

$$U = \frac{Z - \|Z\| \cdot G}{\|Z - \|Z\| \cdot G\|} \quad (2.2.9)$$

where I is the unit matrix. Obviously,

$$H^T H = I$$

$$H^T = H \quad (2.2.10)$$

From (2.2.8) (2.2.10), it can be seen that H is an orthogonal matrix which can rotate vector Z to overlap G . In this problem, V_a and V_1 are unit vectors, so U can be obtained by simply setting $Z = V_a$ and $G = V_1$:

$$U = \frac{V_a - V_1}{\|V_a - V_1\|} \quad (2.2.11)$$

Finally, the transformation matrix T is obtained:

$$T = (T_1, T_2, \dots, T_r) = H(V_1, V_2, \dots, V_r)$$

$$= (V_1^*, V_2^*, \dots, V_r^*) = V^*. \quad (2.2.12)$$

Then,

$$Y = XT = XV^*. \quad (2.2.13)$$

2.3 Further Rotation of V^*

If we consider the RGB colour space case, in which $r=3$, V_1^* , V_2^* and V_3^* determine a 3-D subspace within the m -D feature space. From geometrical point of view, V_1^* , V_2^* and V_3^* also determine a hyperplane P with the normal vector V_1^* (Fig. 1A). For any given V_1^* , V_2^* and V_3^* , they can be rotated around V_1^* by an angle θ to new positions $V_1^*(\theta)$, $V_2^*(\theta)$ and $V_3^*(\theta)$ within P (Fig. 2B). The trace of rotation forms a circle C . It can be seen that $V_1^*(\theta)$, $V_2^*(\theta)$ and $V_3^*(\theta)$ form $V^*(\theta)$. It corresponds to a new projection transformation which satisfies the conditions (2.2.6). When rotating is performed continuously with a given angle increment $\Delta\theta$, a series of transformed data sets $Y(k\Delta\theta)$ is produced:

$$Y(k\Delta\theta) = XV^*(k\Delta\theta), \quad k = 1, 2, \dots \quad (2.3.2)$$

Eq. (2.3.2) will enable interpreters to choose and examine the a data set from various view angle. Because this rotation is on hyperplane P , $V^*(k\Delta\theta)$ can be derived from $V^*(0.0)$ by using basic theorems of the planar geometry.

2.4 Scatter Adjustment of Data Points

In scatter adjustment, every data point is treated as a m -D vector in the feature space. Two nonlinear modes were developed to adjust the scatter. As shown in Fig. 2, V_1 is the central axis, p denotes an arbitrary data point, q is a vector with same direction as V_1 and length equal to $p \cdot q$. The angle and distance between V_1 and p are θ_1 and d_1 ; and θ_2 and d_2 are the angle and distance between V_1 and p^* respectively. The objective of the adjustment is to rotate p to p^* . The amount of the rotation is controlled by the $\theta_2 = k\theta_1$ or $d_2 = kd_1$ according to the mode used. When $k > 1$, this adjustment will result in increase in scatter. When $k < 1$, it will result in decrease in scatter.

In the case of Mode 1, set $d_2 = kd_1$ and let qp^* and p^*p denote the Euclidian distance between the points p and q , p^* and p respectively,

$$\frac{qp^*}{p^*p} = -\frac{k}{k-1} = \lambda, \quad k \neq 1.$$

Then, p^* can be determined,

$$p^* = \frac{q + \lambda p}{1 + \lambda} = kp + (1-k)q \quad (2.4.2)$$

where $q = (V_1 \cdot p)V_1$.

In case of Mode II,

$$\begin{aligned} \text{tg } \theta_1 &= d_1 / \|q\|, \quad \text{tg } \theta_2 = d_2 / \|q\|, \\ \lambda &= -\frac{d_2}{d_2 - d_1} = \frac{d_2}{d_1 - d_2} = \frac{\text{tg } \theta_2}{\text{tg } \theta_1 - \text{tg } \theta_2}. \end{aligned} \quad (2.4.3)$$

Set $\theta_2 = k\theta_1$ then,

$$\lambda = \frac{\text{tg } k\theta_1}{\text{tg } \theta_1 - \text{tg } k\theta_1}, \quad \text{tg } \theta_1 = \frac{\|(v_1 \cdot p)v_1 - p\|}{\|(v_1 \cdot p)v_1\|}.$$

Then

$$p^* = \frac{q + \lambda p}{1 + \lambda}. \quad (2.4.4)$$

To maintain the length of p , (2.4.4) becomes:

$$p^* = \frac{q + \lambda p}{\|q + \lambda p\|} \frac{\|1 + \lambda\|}{1 + \lambda} \|p\| \quad (2.4.5)$$

2.5 The Fisher Ratio Statistical Analysis

As discussed earlier, the results of the transformation will be affected by different rotation angles θ and scatter adjustment factor k . In the following, the Fisher ratio will be tested as a statistical estimation of the effects of the transformation on classification. Let us suppose there are G classes in the data set. The Fisher ratio is defined as the ratio of the between-class scatter S_1 to the within-class scatter S_2 ,

$$\begin{aligned} F &= S_1 / S_2 \\ S_1 &= \sum_{g=1}^G \sum_{l=1}^m \sum_{g'}^n (\bar{x}_{gl} - \bar{x}_1)^2 \\ S_2 &= \sum_{g=1}^G \sum_{k=1}^n \sum_{l=1}^m (x_{gkl} - \bar{x}_{gl})^2 \end{aligned} \quad (2.5.1)$$

Under the projection transformation: $Y = XV^*(\theta)$, F value can be calculated in the following way [4]: first, define the within-class scatter matrix W and between-class scatter matrix B as follows:

$$\begin{aligned} W &= (w_{pq})_{r \times r} \\ w_{pq} &= \sum_{g=1}^G \sum_{k=1}^{n_g} (y_{pgk} - \bar{y}_{pg})(y_{qgk} - \bar{y}_{qg}) \\ B &= (b_{pq})_{r \times r} \\ b_{pq} &= \sum_{g=1}^G n_g (\bar{y}_{pg} - \bar{y}_p)(\bar{y}_{qg} - \bar{y}_q). \end{aligned} \quad (2.5.2)$$

where, $p, q = 1, 2, \dots, r$, and n_g is the number of data points in class g . Then,

$$F = \det B / \det W. \quad (2.5.3)$$

Because Y is the function of rotation angle θ , Eq. (2.5.3) should be written as:

$$F(\theta) = \det B(\theta) / \det W(\theta). \quad (2.5.4)$$

In this research, the between-class scatter matrix is replaced by the total scatter matrix

$$F(\theta) = \frac{\det(B(\theta) + W(\theta))}{\det W(\theta)}. \quad (2.5.5)$$

3. EXPERIMENT AND RESULTS

The method proposed above has been applied to the data sets collected on Sudbury, Ontario. In this experiment, only Quaternary geological problems were considered. According to Boissonneau (1968), there is a generalized glacial history and the experimental area is part of a thinly drift-covered and wave-washed bedrock upland containing isolated pockets of glaciolacustrine and glaciofluvial deposits [7]. The remote sensing data consist of Landsat TM data and other geological and geophysical data set. Six ground classes were selected according to the geological map (Ontario Geological Survey, 1984)(Table 1). They are free of vegetation cover.

3.1 Equal Load of Information

Eigen analysis was carried out on the data set consisting of 6 classes. Some of the results are listed in Table 2. It is clear that there are considerable differences among the variances of the original TM bands and if the first, second and third principal components are displayed in RGB colour space, the percentage of information on each channel will be also very different (Table 3). Therefore, there will be a serious overbalance in RGB space. For comparison, eigen analysis was also performed on the transformed data set Y (Table 4). Table 4 indicates that after the transformation, each new feature has almost equal variance. Because of the existence of correlation among Y_i , equal variance does not necessarily mean that each new feature has an equal amount of information independently. Considering the first variance contribution accounts to 87%, each element of the first eigenvector can be seen as the information load onto the total information from the corresponding Y_i . So, the first eigenvector suggests Y_1, Y_2, Y_3 have approximately the same significance in information space (Table 3). Accordingly, if Y is displayed in RGB colour space, a better balanced resultant image can be expected.

3.2 Information Content after Transformation

Another important question regarding the transformation is how much information can be retained through the transformation. When three bands are selected for RGB colour composite, the information contained in the composite can be measured by the product of three eigenvalues of variance covariance matrix of the three band which represents the volume of the ellipsoid in feature space [1]. Using this method, the amount of information contained in the original 4 bands, the results of first three components of PCA, 3 band composite selected from the original 4 bands and the composite of Y_1, Y_2 and Y_3 are listed in Table 5. From Table 5, the following conclusions may be derived: Firstly, the transformation proposed is as optimal as PCA in terms of maintaining information and very little loss is observed through this transformation. Secondly, any colour composite of three original bands has a considerable information loss. Thirdly, by examining Table 5, the relative information contribution of any of the original TM band can be derived in the following way: Let $I(B)$ stands for information contribution index of band B , then

$$I(5):I(4) = I(7,6,5):I(7,6,4) = 1.26:1$$

$$I(6):I(5) = I(7,6,4):I(7,5,4) = 1.8:1$$

$$I(7):I(6) = I(7,5,4):I(6,5,4) = 1.78:1$$

$$I(4):I(5):I(6):I(7) = 1:1.26:2.26:4.04. \quad (3.2.1)$$

In case of Mode 11,

$$\begin{aligned} \text{tg } \theta_1 &= d_1 / \|q\|, \quad \text{tg } \theta_2 = d_2 / \|q\|, \\ \lambda &= -\frac{d_2}{d_2 - d_1} = \frac{d_2}{d_1 - d_2} = \frac{\text{tg } \theta_2}{\text{tg } \theta_1 - \text{tg } \theta_2}. \end{aligned} \quad (2.4.3)$$

Set $\theta_2 = k\theta_1$ then,

$$\lambda = \frac{\text{tg } k\theta_1}{\text{tg } \theta_1 - \text{tg } k\theta_1}, \quad \text{tg } \theta_1 = \frac{\|(v_1 \cdot p)v_1 - p\|}{\|(v_1 \cdot p)v_1\|}.$$

Then

$$p^* = \frac{q + \lambda p}{1 + \lambda}. \quad (2.4.4)$$

To maintain the length of p , (2.4.4) becomes:

$$p^* = \frac{q + \lambda p}{\|q + \lambda p\|} \frac{\|1 + \lambda\|}{1 + \lambda} \|p\| \quad (2.4.5)$$

2.5 The Fisher Ratio Statistical Analysis

As discussed earlier, the results of the transformation will be affected by different rotation angles θ and scatter adjustment factor k . In the following, the Fisher ratio will be tested as a statistical estimation of the effects of the transformation on classification. Let us suppose there are G classes in the data set. The Fisher ratio is defined as the ratio of the between-class scatter S_1 to the within-class scatter S_2 ,

$$\begin{aligned} F &= S_1 / S_2 \\ S_1 &= \sum_{g=1}^G \sum_{l=1}^n \sum_{k=1}^m (\bar{x}_{gl} - \bar{x}_1)^2 \\ S_2 &= \sum_{g=1}^G \sum_{k=1}^m \sum_{l=1}^n (x_{gkl} - \bar{x}_{gl})^2 \end{aligned} \quad (2.5.1)$$

Under the projection transformation: $Y = XV^*(\theta)$, F value can be calculated in the following way [4]: first, define the within-class scatter matrix W and between-class scatter matrix B as follows:

$$\begin{aligned} W &= (w_{pq})_{r \times r} \\ w_{pq} &= \sum_{g=1}^G \sum_{k=1}^n (y_{pgk} - \bar{y}_{pg})(y_{qgk} - \bar{y}_{qg}) \\ B &= (b_{pq})_{r \times r} \\ b_{pq} &= \sum_{g=1}^G n_g (\bar{y}_{pg} - \bar{y}_p)(\bar{y}_{qg} - \bar{y}_q). \end{aligned} \quad (2.5.2)$$

where, $p, q = 1, 2, \dots, r$, and n_g is the number of data points in class g . Then,

$$F = \det B / \det W. \quad (2.5.3)$$

Because Y is the function of rotation angle θ , Eq. (2.5.3) should be written as:

$$F(\theta) = \det B(\theta) / \det W(\theta). \quad (2.5.4)$$

In this research, the between-class scatter matrix is replaced by the total scatter matrix

$$F(\theta) = \frac{\det(B(\theta) + W(\theta))}{\det W(\theta)}. \quad (2.5.5)$$

3. EXPERIMENT AND RESULTS

The method proposed above has been applied to the data sets collected on Sudbury, Ontario. In this experiment, only Quaternary geological problems were considered. According to Boissonneau (1968), there is a generalized glacial history and the experimental area is part of a thinly drift-covered and wave-washed bedrock upland containing isolated pockets of glaciolacustrine and glaciofluvial deposits [7]. The remote sensing data consist of Landsat TM data and other geological and geophysical data set. Six ground classes were selected according to the geological map (Ontario Geological Survey, 1984) (Table 1). They are free of vegetation cover.

3.1 Equal Load of Information

Eigen analysis was carried out on the data set consisting of 6 classes. Some of the results are listed in Table 2. It is clear that there are considerable differences among the variances of the original TM bands and if the first, second and third principal components are displayed in RGB colour space, the percentage of information on each channel will be also very different (Table 3). Therefore, there will be a serious overbalance in RGB space. For comparison, eigen analysis was also performed on the transformed data set Y (Table 4). Table 4 indicates that after the transformation, each new feature has almost equal variance. Because of the existence of correlation among Y_i , equal variance does not necessarily mean that each new feature has an equal amount of information independently. Considering the first variance contribution accounts to 87%, each element of the first eigenvector can be seen as the information load onto the total information from the corresponding Y_i . So, the first eigenvector suggests Y_1, Y_2, Y_3 have approximately the same significance in information space (Table 3). Accordingly, if Y is displayed in RGB colour space, a better balanced resultant image can be expected.

3.2 Information Content after Transformation

Another important question regarding the transformation is how much information can be retained through the transformation. When three bands are selected for RGB colour composite, the information contained in the composite can be measured by the product of three eigenvalues of variance covariance matrix of the three band which represents the volume of the ellipsoid in feature space [1]. Using this method, the amount of information contained in the original 4 bands, the results of first three components of PCA, 3 band composite selected from the original 4 bands and the composite of Y_1, Y_2 and Y_3 are listed in Table 5. From Table 5, the following conclusions may be derived: Firstly, the transformation proposed is as optimal as PCA in terms of maintaining information and very little loss is observed through this transformation. Secondly, any colour composite of three original bands has a considerable information loss. Thirdly, by examining Table 5, the relative information contribution of any of the original TM band can be derived in the following way: Let $I(B)$ stands for information contribution index of band B , then

$$\begin{aligned} I(5):I(4) &= I(7,6,5):I(7,6,4) = 1.26:1 \\ I(6):I(5) &= I(7,6,4):I(7,5,4) = 1.8:1 \\ I(7):I(6) &= I(7,5,4):I(6,5,4) = 1.78:1 \\ I(4):I(5):I(6):I(7) &= 1:1.26:2.26:4.04. \end{aligned} \quad (3.2.1)$$

[7] Burwasser G.J., 1979. Ontario Geological Survey Report 181: Quaternary Geology of the Sudbury Basin Area, District of Sudbury, pp 10-11.

[8] Ministry of Natural Resources, Ontario, 1984. The Geology and Ore Deposits of the Sudbury Structure, Special Vol. No. 1, pp 57-82, Government of Ontario.

FIGURE CAPTIONS

1.A: The projection transformation with equal information distribution.

1.B: Rotating V^* to $V^*(\Delta\theta)$.

2: Adjusting scatter by rotating data vectors.

3: Mean vectors for class 1, 2, and 3.

4: The Fisher ratio curves with different parameters.

TABLE 1

| Class No. | Number of Samples | Geological Formation |
|-----------|-------------------|---|
| 1 | 81 | Onaping Formation, a variety of heterolithic breccias |
| 2 | 63 | Onwatin Formation, carbonaceous and pyritic mudstone |
| 3 | 48 | Chelmsford Formation, wacke |
| 4 | 45 | Sudbury Igneous Complex, granophyre |
| 5 | 63 | Sudbury Igneous Complex, norite |
| 6 | 55 | Unsubdivided granitic rocks |

TABLE 2

| Eigenvalue No. | Eigenvalue | Variance Contribution (%) |
|----------------|------------|---------------------------|
| 1 | 45.15 | 85.48 |
| 2 | 3.69 | 6.99 |
| 3 | 2.61 | 4.94 |
| 4 | 1.37 | 2.59 |

TABLE 3

| TM Band No. | 7 | 6 | 5 | 4 |
|-------------------|-------|-------|------|------|
| Variance | 25.63 | 16.14 | 7.95 | 3.1 |
| First Eigenvector | 0.73 | 0.56 | 0.36 | 0.16 |

TABLE 4

| New Feature No. | 1 | 2 | 3 |
|-------------------|-------|-------|-------|
| Variance | 17.46 | 16.84 | 17.15 |
| First Eigenvector | 0.57 | 0.49 | 0.45 |

TABLE 5

| Data Set | TM | PCA | New Feature | TM | TM | TM | TM |
|------------------------------|---------|------|-------------|-------|-------|-------|-------|
| Selected Bands | 7,6,5,4 | 123 | 123 | 7,6,5 | 7,6,4 | 7,5,4 | 6,5,4 |
| Information Index | 594 | 434 | 434 | 337 | 267 | 148 | 84 |
| The Ratio to Index (7,6,5,4) | 1 | 0.73 | 0.73 | 0.57 | 0.45 | 0.24 | 0.14 |

TABLE 6.1 $V^*(0)$

| Matrix |
|---------------------|
| 0.083 -0.462 -0.882 |
| -0.570 -0.737 0.338 |
| -0.677 0.296 -0.247 |
| -0.458 0.395 -0.216 |

Table 6.2 $V^*(45)$

| Matrix |
|---------------------|
| 0.107 -0.844 -0.524 |
| -0.936 -0.245 0.213 |
| -0.319 0.324 -0.634 |
| -0.102 0.351 -0.528 |

TABLE 6.3

| Feature No. | 1 | 2 | 3 |
|-------------------|------|------|------|
| Variance | 17.5 | 16.8 | 17.1 |
| First Eigenvector | 0.58 | 0.49 | 0.45 |

TABLE 6.4

| Feature No. | 1 | 2 | 3 |
|-------------------|------|------|------|
| Variance | 17.3 | 17.3 | 16.8 |
| First Eigenvector | .58 | .49 | .45 |

TABLE 6.5

| Feature No. | 1 | 2 | 3 |
|-------------|------|------|-----|
| 1 | 1.0 | | |
| 2 | .816 | 1.0 | |
| 3 | .791 | .842 | 1.0 |

TABLE 6.6

| Feature No. | 1 | 2 | 3 |
|-------------|------|------|-----|
| 1 | 1.0 | | |
| 2 | .787 | 1.0 | |
| 3 | .831 | .831 | 1.0 |

TABLE 7 R Matrix with Different Scatter

| | Mode I | | | Mode II | | |
|-------------|--------|------|---|---------|------|---|
| Feature No. | 1 | 2 | 3 | 1 | 2 | 3 |
| k=1.5 | 1 | | | 1 | | |
| 2 | .589 | 1 | | .646 | 1 | |
| 3 | .669 | .669 | 1 | .702 | .723 | 1 |
| k=2.0 | 1 | | | 1 | | |
| 2 | .389 | 1 | | .551 | 1 | |
| 3 | .502 | .502 | 1 | .602 | .656 | 1 |

TABLE 8 Between-Class Distance Matrix

| X Data | | | | Y Data | | | |
|-----------|------|------|---|-----------|------|------|---|
| Class No. | 1 | 2 | 3 | Class No. | 1 | 2 | 3 |
| 1 | 0 | | | 1 | 0 | | |
| 2 | 4.09 | 0 | | 2 | 6.42 | 0 | |
| 3 | 1.51 | 2.59 | 0 | 3 | 2.44 | 4.05 | 0 |

A METHOD FOR MULTI-DIMENSIONAL IMAGE SEGMENTATION

Keith Hartt, Mark J. Carlotto, and Mark W. Brennan

The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, Massachusetts 01867
(617)942-2000

ABSTRACT

A multi-dimensional (multi-spectral) image segmentation technique is presented, based on a multivariate Gaussian mixture distribution which includes spatial dependence through the incorporation of a Markov random field that governs the grouping of pixel classifications into regions. Estimation of parameters for each component in the mixture, an unsupervised clustering problem, is performed using a specialization of the EM algorithm, an iterative scheme for maximum likelihood parameter estimation. The commonly used heuristic Isodata algorithm is cast in this theoretical framework and is shown to be an approximate method under a restricted model assumption; furthermore, a computational approach is presented which allows our more general procedure to be employed with comparable efficiency. As the EM algorithm is essentially a parameter-refinement procedure, which only locally maximizes the likelihood function, good initial estimates for parameters are crucial in obtaining satisfactory results. Two methods are presented in this regard which involve incremental building of model complexity: 1) a histogram analysis technique in which a smoothed 1D cluster histogram is parsed into a number of components based on inflection points; 2) a straightforward method in which a maximum-extent cluster is replaced by a two-component mixture. The segmentation estimate is optimal in the sense that, based on the statistical model, it is the most probable outcome for pixel classifications given the multi-dimensional imagery. Our segmentation technique is being used extensively in the building of a surface-material-class database from Landsat TM imagery; practical experience with the method is presented.

1. INTRODUCTION

Methods for multi-dimensional image segmentation abound; surveys of previous work can be found in [7], [9], [12]. The goal of image segmentation is to extract a set of regions in which certain important characteristics are homogeneous. Kanade [12] divides multispectral image segmentation techniques into three categories comprised of: 1) those which use local spatial information for region merging, 2) those which use spectral information (i.e., the spatially independent distribution of the multi-dimensional data) for region splitting; 3) those which use spectral and spatial information.

In this paper a technique is presented with which we have had practical success, and which has a computationally efficient implementation. It is based on computational techniques and a realistic image model with sound theoretical underpinnings. The method is a member of category 3), as it is based on a mixture distribution which incorporates spatial dependence. As is customary, each component, or cluster, in the mixture is modeled using a normal distribution. Here, an unsupervised clustering technique is used to estimate the number of clusters and the statistical characteristics of each cluster. One method presented for parsing, or sub-dividing, clusters is a scale-space histogram analysis technique [3] extended to a multi-dimensional setting. Cluster parameters are refined using an iterative technique. The spatial grouping of pixels into regions, i.e., the classification of each pixel as a member of a cluster, is governed by a categorically valued Markov random field which has a multi-level logistic distribution [6]. Image models using Markov random fields for segmentation and restoration in a single-image setting have been widely used [5], [6], [9], [13]. The model promotes the formation of compact regions by assigning a high probability to groupings of identically classified pixels. The optimization criterion is based on a Bayesian formulation; in

particular, the segmentation is a MAP estimate, the most probable pixel classifications given the multi-dimensional data. In the computation of the optimal segmentation, a relaxation procedure is used which involves locally updating the segmentation estimate [5], [9].

The paper is organized as follows. In Section 2, the clustering procedure is described. The output of the clustering is essentially parameters used in the segmentation, which is described in Section 3. Experimental results are given in Section 4.

2. CLUSTERING

The model used for clustering is given by the following normal mixture distribution, which is the marginal distribution at all pixel sites of the multi-dimensional image data.

$$p(x) = \sum_k \pi_k \left(\frac{1}{\prod_m \sqrt{2\pi} \sigma_{m,k}} \right) \exp \left(-\frac{1}{2} \sum_m \left(\frac{x_m - \mu_{m,k}}{\sigma_{m,k}} \right)^2 \right) \quad (1)$$

where m indexes data dimension, x denotes $\{x_m: \text{all } m\}$, k indexes the cluster or component; π_k is a relative frequency or prior weight, where $\sum \pi_k = 1$; $\mu_{m,k}$ is a mean, $\sigma_{m,k}$ a standard deviation. Note that the i.i.d. assumption relative to location in the lattice is relaxed in the segmentation technique presented later.

Some words justifying the clustering model are in order. The general aim is to represent the data using a probability distribution comprised of unimodal distributions, each of which corresponds to a cluster of data. The data are described compactly by cluster membership. A normal distribution provides a tractable, reasonable unimodal distribution. Note that if the number of clusters is not upper-bounded, the model is perfect in the sense that each multi-dimensional datum can be assigned an infinitesimally narrow unimodal distribution centered on itself, resulting in the data having a likelihood of one. This of course is not a useful model, so the number of components must be reasonably restricted. The model assumes a diagonal covariance matrix for component normal distributions for simplicity; the computational burden is relaxed considerably relative to a general covariance matrix. Furthermore, even if an independence assumption is not valid, the model is adequate if the marginal description affords inter-cluster statistical separability. In our view, based on experiment, a compute-time/generality tradeoff supports the model used here.

EM Algorithm

The EM algorithm is a general procedure for maximum likelihood estimation given incomplete data. The specialization of the EM algorithm applicable to the mixture estimation problem at hand is given by the following iteration [16]

$$\begin{aligned} \hat{\pi}_k^{(n)} &= \frac{1}{N^2} \sum_{i,j} p^{(n-1)}(k | x_{ij}) \\ \hat{\mu}_{m,k}^{(n)} &= \frac{1}{\hat{\pi}_k^{(n)} N^2} \sum_{i,j} x_{m,i,j} p^{(n-1)}(k | x_{ij}) \\ \hat{\sigma}_{m,k}^{(n)} &= \frac{1}{\hat{\pi}_k^{(n)} N^2} \sum_{i,j} (x_{m,i,j} - \hat{\mu}_{m,k}^{(n)})^2 p^{(n-1)}(k | x_{ij}) \end{aligned} \quad (2)$$

where n indexes the iteration, (i, j) indexes location in the image lattice consisting of N^2 pixels, $x_{m,ij}$ denotes $\{x_{m,ij}; \text{all } m\}$. Parameter estimates for each cluster are in the form of sample statistics weighted by a cluster membership probability, given by

$$p^{(n-1)}(k | x_{ij}) = \frac{p^{(n-1)}(x_{ij} | k) \pi_k}{\sum_k p^{(n-1)}(x_{ij} | k) \pi_k} \quad (3)$$

where $p^{(n-1)}(x_{ij} | k)$ is the normal distribution of the multi-dimensional data for component k , computed using the parameter set $\{\pi_k, \mu_{m,k}, \sigma_{m,k}\}^{(n-1)}$ obtained at iteration $(n-1)$. Note k is interpreted as a random integer as well as an index to clarify the presentation; e.g., $p(k)$ could be used to denote π_k . The convergence behavior of the EM algorithm is that at each iteration, the likelihood function is increased, more information in this regard can be found in [16]. The measure used to test for convergence of the EM algorithm is given by the discrimination information [14] between mixture distributions at successive iterations

$$I = \sum_m \sum_k \frac{1}{2} (\sigma_{m,k}^{2(n)} - \sigma_{m,k}^{2(n-1)}) (\sigma_{m,k}^{2(n)} - \sigma_{m,k}^{2(n-1)}) + \frac{1}{2} (\sigma_{m,k}^{2(n)} + \sigma_{m,k}^{2(n-1)}) (\mu_{m,k}^{(n)} - \mu_{m,k}^{(n-1)})^2 \quad (4)$$

Discrimination information between the model at one iteration and the previous one falling below a specified threshold signals convergence.

It is worthwhile to note the relationship of this algorithm to the commonly used Isodata algorithm [7]. The mixture EM algorithm is equivalent under the assumption that standard deviations and relative frequencies are uniform across clusters, and when the cluster membership probability is approximated by

$$p(k | x_{ij}) = 1 \text{ if } k = \arg \max_k p(k | x_{ij}) \\ 0 \text{ otherwise} \quad (5)$$

In essence, a hard decision about cluster membership is made at each iteration by using the approximation, under the assumption that relative frequencies and standard deviations are uniform, this decision at each pixel site is given by the cluster with its center closest in the Euclidean sense to the multi-dimensional datum. The members of each cluster then each contribute to a new estimate for their center to be used in the next iteration.

Cluster Parsing

The EM algorithm is a parameter refinement procedure. Initial estimates of parameter values are necessary, and the number of components must be specified. In this section, procedures for building up the model by subdividing clusters is presented. The methods involve first assuming a single cluster, and then recursively subdividing.

A simple heuristic for subdividing is to divide in two the cluster having the largest standard deviation, in the hope of obtaining a model consisting of compact clusters. The parameters of the subdivided cluster are related to the parent cluster by

$$\pi_0 = \pi_1 = \frac{\pi}{2}; \mu_0 = \mu - \frac{\sigma}{\sqrt{2}}; \mu_1 = \mu + \frac{\sigma}{\sqrt{2}}; \sigma_0 = \sigma_1 = \frac{\sigma}{\sqrt{2}} \quad (6)$$

where $\{\pi, \mu, \sigma\}$ is the parameter set associated with cluster k along dimension m , with

$$\{m, k\} = \arg \max_{\{m, k\}} \sigma_{m,k} \quad (7)$$

All other parameters remain unchanged. The relationship (6) is such that the two-component mixture distribution is equivalent to the parent distribution up to the second moment. This sort of heuristic is commonly used in conjunction with Isodata [17]. It is more attractive here however, as standard deviations are an explicit part of the clustering model.

A second method for cluster sub-division is based on a multi-dimensional extension of scale-space histogram analysis [3]. The method (and implementation) is based on the following statistic, which is an empirical marginal distribution, a cluster histogram for one

dimension of the multi-dimensional data.

$$h_k(x_m) = \frac{1}{N^2} \sum_{\{(i,j): x_{m,ij} = x_m\}} p(k | x_{ij}) \quad (8)$$

Note that this is a sufficient statistic for calculations in (2). Compute time is reduced because once the statistic is computed, the large sums over all (i, j) are replaced by small sums over all values of x_m , for typical eight-bit images. Use of the statistic lessens the computation carried out at each pixel site, the computation which dominates overall compute time. Additionally, in parsing clusters, the histogram statistic allows the use of scale-space histogram analysis in a multi-dimensional setting.

Scale-space histogram analysis is described briefly as follows. The method performs a modal analysis on a 1D histogram using inflection points, zero-crossings of the second derivative, in a smoothed version of the histogram. Initial estimates for cluster parameters are related to the inflection points $\{x_n\}$ by

$$\hat{\pi}_n = \int_{x_n}^{x_{n+1}} h_s(x) dx \\ \hat{\mu}_n = \frac{1}{2} (x_n + x_{n+1}); \hat{\sigma}_n = \frac{1}{2} (x_{n+1} - x_n) \quad (9)$$

for n even, where $h_s(x)$ is the smoothed histogram, and relative frequencies are appropriately normalized. A Gaussian-shaped smoothing kernel is used, because it has the attractive property that the number of modes monotonically decreases with increasing kernel width [3].

The method is applied to the multi-dimensional data by applying the technique recursively to the 1D cluster histograms. The multi-dimensional data are weighted by the membership probability for a cluster, and the weighted data are then histogrammed along a dimension of the multi-dimensional data, resulting in a cluster histogram to which histogram analysis is applied. If multiple modes are found, the cluster is subdivided according to (9) for that dimension; all other parameters remain unchanged. The procedure is repeated, each time followed by EM iteration, for all clusters along all dimensions until all cluster histograms are unimodal by the histogram analysis criterion. The recursive process results in a powerful yet manageable method for the parsing of multi-dimensional clusters. An alternate interpretation of the formation of the cluster histograms is as follows. The multi-dimensional data histogram bins are weighted by cluster membership probabilities. The result is an empirical distribution of the cluster of multi-dimensional data. The weighted bins are projected onto coordinate axes, forming empirical marginal distributions.

Computational Considerations

The computational impact of using the cluster histogram statistic has been described. In addition, a look-up table approach is taken for the calculation of cluster-membership probabilities, which are given by

$$p(k | x_{ij}) = \frac{a_k}{\sum_k a_k} \quad (10) \\ a_k = \exp \left(\log(\pi_k) - \sum_m \left(\log(\sigma_{m,k}) + \frac{1}{2} \left(\frac{x_{m,ij} - \mu_{m,k}}{\sigma_{m,k}} \right)^2 \right) \right)$$

The elements of the sum in the exponent are precomputed before sweeping the image, for each possible value of $x_{m,ij}$. For typical eight-bit images, the computations comprise a small set, which allows storage in a look-up table. With the remaining logarithm in a_k precomputed, the computation of the cluster-membership probability at each pixel site involves a number of sums, an exponentiation, and a division. The result is that the total amount of computation is comparable to that in Isodata, which makes our algorithm an attractive, more general alternative.

3. SEGMENTATION

The result of clustering is a set of parameters for each cluster in the mixture model. In this section the segmentation, the final hard decision about the cluster membership of each multi-dimensional datum on the image lattice, is presented. The model used for segmentation is identical

to that used for clustering with the exception that the independence assumption associated with marginal description of the cluster membership of each pixel is relaxed. Toward this end, the enhanced model for cluster membership which incorporates spatial dependence is described. The optimization criterion is then derived based on the model, and computational aspects of the segmentation algorithm are presented.

Multi-Level Logistic Model

The multi-level logistic distribution was proposed by Derin and Elliott in their work on single-image segmentation [6]. The distribution is conveniently presented in the context of Markov random fields, described through Gibbs distributions. The main utility of the Gibbs distribution lies in the fact that a random field can be described in a tractable manner in terms of local interactions which are in accord with desired behavior. After some preliminaries on general Gibbs models, the specifics of the multi-level logistic model are provided.

The Markov property on a lattice is defined with respect to a neighborhood system $\{n_{ij}\}$, where n_{ij} denotes the neighborhood of pixel site (i,j) . The Markov property is given by

$$p(w_{ij} | \{w_{kl} : (k,l) \neq (i,j)\}) = p(w_{ij} | \{w_{kl} : (k,l) \in n_{ij}\}) \quad (11)$$

where $\{w_{ij}\}$ is a realization of a general 2D field. A Markov random field can always be expressed in the form of a Gibbs distribution [1]

$$p(\{w_{ij}\}) = \frac{1}{Z} \exp(-\sum_{c \in C} V_c(\{w_{ij}\})) \quad (12)$$

where c is a set of pixels that are neighbors of each other, called a *clique*; C is the set of all cliques; $V_c(\{w_{ij}\})$ is a *clique potential* which describes the interaction among members of clique c ; Z is a normalizing constant. The relationship between the Gibbs distribution and the Markov property is demonstrated by the following. Consider

$$p(w_{ij} | \{w_{kl} : (k,l) \neq (i,j)\}) = \frac{p(\{w_{ij}\})}{\int p(\{w_{ij}\}) dw_{ij}} \quad (13)$$

where $p(\{w_{ij}\})$ is given by (12). Note that the normalizing constant and all clique potentials cancel with the exception of the potentials involving cliques of which (i,j) is a member. This gives rise to the Markov property (11).

The multi-level logistic distribution used here describes the cluster-membership field, consisting of a set of cluster memberships, one for each pixel site, denoted by $\{k_{ij}\}$; i.e., it is a 2D field of random integers which label clusters. Clique potentials are given by

$$V_c(\{k_{ij}\}) = -\omega_c \quad \text{if all } k_{ij} \text{ in } c \text{ are equal} \\ \omega_c \quad \text{otherwise} \quad (14)$$

Only certain pair clique potentials are non-zero, which results in a field Markov with respect to the eight nearest neighbors. Cliques and associated potentials are depicted in Fig. 1.

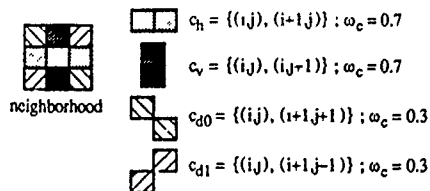


Fig. 1. Clique potentials.

Optimization Criterion

The segmentation is based on the MAP estimate, given by

$$\hat{k}_{ij} = \arg \max_{\{k_{ij}\}} p(\{k_{ij}\} | \{x_{ij}\}) \quad (15)$$

which is the most probable set of cluster memberships given the set of multi-dimensional data. The posterior distribution is given by Bayes'

rule

$$p(\{k_{ij}\} | \{x_{ij}\}) = \frac{\prod_{ij} p(x_{ij} | k_{ij}) p(\{k_{ij}\})}{p(\{x_{ij}\})} \quad (16)$$

Note that as in the clustering model, the multi-dimensional data are assumed independent given cluster membership, so that the joint conditional distribution is expressed as a product of marginals. However, the cluster-membership field incorporates dependence through the multi-level logistic model of spatial interaction. Spatial as well as spectral (in the case of multi-spectral data) clustering is incorporated in the segmentation model. The choice of clique potentials above associates a high probability with spatially compact regions consisting of identically classified pixel sites. Note that the denominator in (16) has no dependence on cluster membership, so is simply a normalizing constant which does not effect the maximization (15).

Relaxation Method

Direct computation of the maximization (15) is not practically feasible, as it involves a search over all possible values of the cluster-membership field, which grows exponentially with N^2 . A relaxation method is therefore presented which involves iteratively updating the cluster membership at each site. The update involves a local computation given by the following conditional distribution derived from (16)

$$p(k_{ij} | \{k_{kl} : (k,l) \neq (i,j)\}, \{x_{ij}\}) = \frac{p(x_{ij} | k_{ij}) p(\{k_{ij}\})}{\sum_{k_{ij}} p(x_{ij} | k_{ij}) p(\{k_{ij}\})} \quad (17)$$

where the sum in the denominator is over all values of k_{ij} . Note that the posterior distribution is Markov, so that the computation involves only the multi-dimensional datum and the cluster membership at site (i,j) , and its neighborhood of cluster memberships. The calculation for the particular multi-level logistic model used here involving only pair cliques is given by

$$p(k_{ij} | \{k_{kl} : (i,j) \in n_{ij}\}, x_{ij}) = \frac{a(k_{ij})}{\sum_{k_{ij}} a(k_{ij})} \quad (18)$$

$$a(k_{ij}) = \exp \left(\sum_{\{c: (i,j) \in c\}} 2\omega_c \# \{(\kappa,l) : (\kappa,l) \neq (i,j), (\kappa,l) \in c, k_{\kappa l} = k_{ij}\} - \sum_m \left(\log(\sigma_{m,k_{ij}}) + \frac{1}{2} \left(\frac{x_{m,i,j} - \mu_{m,k_{ij}}}{\sigma_{m,k_{ij}}} \right)^2 \right) \right)$$

The elements in the first summation in the exponent are either $2\omega_c$ or zero, depending on whether the membership neighboring (i,j) in the clique is equal to k_{ij} or not, respectively. This results in a conditional probability of an outcome k_{ij} which increases with the number of neighbors having membership k_{ij} . It is instructive to compare this expression with (10), the clustering-model membership probability distribution. It is identical with the exception of the first term in the exponent, associated with the distribution of cluster memberships which, because of spatial dependence, are indexed relative to the image lattice. The look-up table method of computation is employed here as well.

The relaxation method is a deterministic version of the Gibbs sampler [9]. The image is repeatedly swept, with replacement at each pixel site given by

$$\hat{k}_{ij} = \arg \max_{k_{ij}} p(k_{ij} | \{k_{kl} : (i,j) \in n_{ij}\}, x_{ij}) \quad (19)$$

the estimate which is the locally most probable value. Following the relaxation procedure, the state of the cluster membership field converges to a local maximum of the *global* posterior distribution (16). To avoid imposing the time causality on the lattice, the site visit ordering is chosen according to the coding method [1]. The initial configuration is the maximum likelihood estimate under an independence assumption, which is given by (19) with the conditional probability (18) modified by setting the first term in the exponent to zero.

4. EXPERIMENTAL RESULTS

Practical experience with the method is demonstrated with a surface-material-classification application to six-band Landsat Thematic Mapper (TM) multi-spectral data. Our method is currently being used in this way to produce a database over a large geographical area.

Here, a (1383×1125)-pixel area was extracted from TM Path 14 Row 30 which covers a study area surrounding Glens Falls, NY. The tasseled-cap transformation [4] was applied to the image resulting in brightness, greenness, and wetness images. This transformation is a linear rotation which allows reducing the dimensionality of the data in a manner appropriate for the sensor and the discrimination of natural surface materials. Figure 2 shows the brightness image, which is essentially panchromatic. A hybrid approach to image classification combining supervised and unsupervised classification strategies was taken. The approach involves the manual selection of training sets for each surface material category. However, the distribution of each category is not assumed unimodal; i.e., the clustering algorithm is applied to each category. The clusters for each category are combined, and the segmentation algorithm is applied. Figure 3 shows the segmentation result, where the grey-level mapping is given by the cluster mean in brightness. A total of 51 clusters are represented, each of which corresponds to one of nine surface material classes of interest: agriculture, urban, bare soil, open water, grassland, brush/scrub, deciduous forest, coniferous forest, and mixed forest.

Accuracy of the surface material classification obtained using the segmentation was measured based on ground truth obtained from 1:15,840 scale SCS orthophotos, field notes and terrestrial photographs. A 95% confidence interval was constructed for the overall classification accuracy according to [15]. The overall classification accuracy with respect to the nine surface material classes was 86.8%, which is good relative to other research using similar classification categories [11], [18]. The agriculture, open water, grassland, coniferous forest and mixed forest classes yielded the best results (82.9 - 95.6 %). The lowest classification agreement was for the urban (79.5 %) and brush/scrub (58.5 %) classes, which is attributable to their consisting of relatively complex mixtures. More detail on this experiment may be found in [2].

Summary

A new segmentation algorithm is described using a rigorous statistical framework. The presentation includes the extension of a histogram analysis technique to a multi-dimensional setting, as well as the extension of results of single-image segmentation studies for application to multi-dimensional data.

REFERENCES

- [1] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. Ser. B*, vol. 36, 1974, pp. 192-236.
- [2] M. W. Brennan, T. J. Moore, and B. L. Spence, "Ten meter database development using Landsat Thematic Mapper and digitized map data: preliminary results," *Tech. Papers 1989 ASPRS-ACSM Ann. Conv.*, vol. 3, pp. 165-172.
- [3] M. J. Carlotto, "Histogram analysis using a scale-space approach," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-9, no. 1, Jan. 1987, pp. 121-129.
- [4] E. P. Crist and R. C. Ciccone, "A physically-based transformation of Thematic Mapper data -- the TM tasseled cap," *IEEE Trans. Geo. Rem. Sens.*, vol. GE-22, no. 3, May 1984, pp. 256-263.
- [5] F. S. Cohen and D. B. Cooper, "Real time textured image segmentation based on noncausal Markovian random field models," *Proc. SPIE Conf. Intel. Robots*, Cambridge, MA, Nov. 1983.
- [6] H. Denn and H. Elliou, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-9, no. 1, Jan. 1987, pp. 39-55.
- [7] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.
- [8] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pat. Recog.*, vol. 13, pp. 3-16.
- [9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-6, Nov. 1984, pp. 721-741.
- [10] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comp. Vis., Gr., Img. Proc.*, vol. 29, 1985, pp. 100-132.
- [11] M. Hashim, "Crop identification using merged Landsat multispectral scanner and Thematic Mapper data: preliminary analysis," *Tech. Papers 1988 ASPRS-ACSM Ann. Conv.*, vol. 4, pp. 11-20.
- [12] T. Kanade, "Region segmentation: signal vs semantics," *Computer Graphics and Image Processing*, vol. 19, 1980, pp. 279-297.
- [13] P. A. Kelly, H. Derin and K. D. Hart, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-36, Oct. 1988, pp. 1628-1641.
- [14] S. Kullback, *Information Theory and Statistics*, New York: Dover, 1968.
- [15] J. Neter, W. Wasserman, and G. A. Whitmore, *Applied Statistics*, Boston: Allyn and Bacon, Inc., 1982.
- [16] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, Apr. 1984, pp. 195-239.
- [17] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Reading, MA: Addison Wesley, 1974.
- [18] J. Vogel, "An evaluation of the use of TM digital data for updating the land cover component of the SCS 1987 multiresource inventory of New Jersey," *Tech. Papers 1987 ASPRS-ACSM Ann. Conv.*, vol. 1, pp. 113-119.

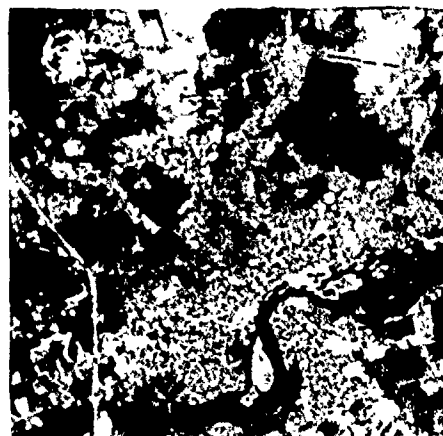


Fig. 2. Brightness image.



Fig. 3. Segmentation.

AN ADAPTIVE RECONSTRUCTION SYSTEM FOR SPATIALLY CORRELATED MULTISPECTRAL MULTITEMPORAL IMAGES

Lee, S. and Crawford, M. M.

Department of Mechanical Engineering
and
UT Center for Space Research
The University of Texas at Austin, TX, USA

Abstract.

An adaptive reconstruction system has been developed to analyze sequential images observed at regular time intervals. A least-squares linear predictor with escalator structure has been implemented in the new system. Using the predictor, estimates of missing data or bad (possible cloud covered) data and spatial parameters at a specified time can be obtained from previous history. This algorithm recovers from observations which are contaminated due to blurring and correlated noise using temporally adapted spatial parameters.

Keywords: image reconstruction; adaptive system; spatial parameters; Bayesian estimation; multispectral time series.

1. Introduction.

During the last decade, a wide range of statistical techniques have been developed for analyzing multispectral images in spatial context. However, statistical approaches to multitemporal analysis of images remain largely unexplored, although the temporal component contains abundant, useful information. The evolution of technology is radically affecting the quantity and quality of data collected through remote sensing. It is now possible to continuously acquire images at regular time intervals. Multitemporal features have usually been exploited through artificial intelligence approaches. Recently, there has been increased interest in the use of statistical multitemporal techniques in a spatial context. Khazenie (1987) developed a contextual classification method which incorporates constant temporal correlation.

In multitemporal image analysis, there is a high likelihood that during at least one of the data acquisition periods the target site corresponding to any given pixel may be covered by clouds, thereby resulting in missing data. Of even greater importance is the need to incorporate temporal variation of the spectral component according to physical properties of targets and atmospheric changes into image processing techniques. These problems can be overcome by "automatic unsupervised learning" in the reconstruction system through an adaptive predictor. The new predictor, which provides estimates of missing observations and of adaptive spatial parameters, uses an escalator structure associated with "Gram-Schmidt orthogonalization" (Ahmed and Youn, 1980). With this approach, it is possible to analyze a sequence of images as an on-line process.

In digital remote sensing, the original distribution of radiated intensity is modified by residual effects (blurring) re-

sulting from imperfect sensing of target and by spatially auto-correlated noise due to atmospheric attenuation of the signal. Image reconstruction algorithms are based on the premise that a reasonable representation of the original image can be recovered from a blurred, noisy version of the true target value. The effectiveness of the algorithms depends to a great extent on the validity of the image model. A statistical model based on a general structure of the digital image process is proposed as one component of the reconstruction system to improve pictorial information for image interpretation.

The original uncontaminated intensity which is one characteristic of the physical processes of targets can account for random variation about the mean intensity that is inherent in the target characteristics. The new adaptive image reconstruction procedure is based on a multiple compound process: (a) the mean intensity process is a Gibbs random field (GRF) (Georgii, 1979), (b) the original intensity process is a multispectral time series (Robinson, 1967) and (c) the observation process is a spatial-autoregressive model (Ripley, 1981). The feedback system, which combines three filters, is outlined in Figure 1. The algorithm can be summarized as follows. Given an image observation and a probability structure of the process, the original intensity is restored in the Bayesian restoration filter. The linear predictor yields the least-squares prediction and the covariance structure of the original intensity for the next stage. The prediction is used to estimate the spatial parameters and the mean intensity, which are required for the Bayesian restoration filter. The linear predictor easily detects temporal changes in the spectral behavior of sequential scenes, which is of primary interest in many applications.

2. Digital Image Model.

The following assumptions, which can properly describe fundamental nature of actual physical processes, are used for modelling the image process: (a) the original intensity process is represented by a discrete multispectral time series, (b) a GRF characterizes the mean intensity process, (c) the distribution of original intensity is acted upon by homogeneous (shift-invariant) linear systems to produce a spatially correlated image, (d) the spatial-noise is identical over the image space, (e) the temporal-noise is signal-dependent and correlated across spectral bands and (f) the noise is additive and Gaussian. It is assumed that the spatial dependency across

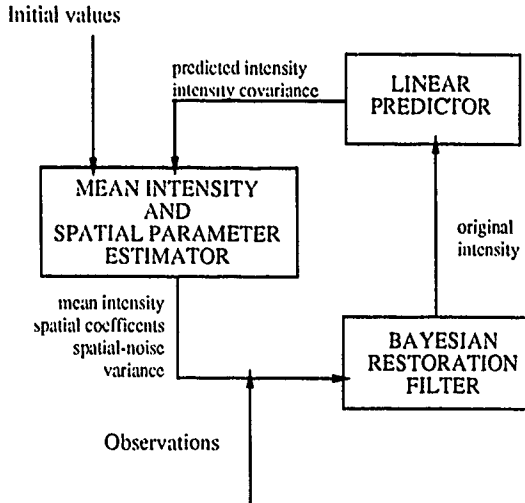


Figure 1. Adaptive reconstruction system.

time exhibited in the physical behavior of targets is assimilated into the spatial component resulting from the residual effects and the GRF textured model.

Consider a sample image of d spectral bands with a set of indices of the pixels $I_n = \{1, \dots, n\}$. If $Y_{(t),i}$, $X_{(t),i}$ and $\mu_{(t),i}$ are vectors denoting the observation, the original intensity and the mean intensity for the i th pixel at time t , for $i = 1, \dots, n$, the image process is modeled as,

$$Y_{(t),i} = U_{(t),i} + \sum_{j \in I_n \setminus i} A_{ij} [Y_{(t),j} - U_{(t),j}] + \varepsilon_{(t),i}$$

$$Z_{(t),i} = \sum_{j=1}^p C_{ij}^z Z_{(t-j),i} + \eta_{(t),i} + \sum_{j=1}^q C_{ij}^\eta \eta_{(t-j),i}$$

$$U_{(t),i} = \sum_{j \in I_n} B_{ij} X_{(t),j}, \quad Z_{(t),i} = X_{(t),i} - \mu_{(t),i}$$

where A_{ij} and B_{ij} are the spatial coefficients associated with the i th and j th pixels. C_{ij}^z and C_{ij}^η are the autoregressive and moving-average temporal coefficients associated with the i th and j th time-lags. $\varepsilon_{(t)}$ and $\{\eta_{(t),i}\}$ are independent Gaussian noise vectors. The $d \times d$ matrices of spatial coefficients are diagonal. Temporal coefficients are contained in general $d \times d$ matrices. If the spatial coefficients $A = \{A_{ij}\}$ are proportional to the relative proximity of the pixel pair, then $A = I - \gamma D$. The proximity matrix D , whose diagonal elements are all zero, is operated on by the diagonal matrix γ which has n blocks identical of d elements on the diagonal. The elements of ρ corresponding to the identical block determine the level and direction of spatial-noise autocorrelation in the corresponding spectral bands. It is common practice to use this assumption to simplify the model for mathematical tractability. Block-rows of the blurring operator $B = \{B_{ij}\}$ and the proximity matrix $D = \{D_{ij}\}$ are related by a circular-shift in a homogenous linear system if the boundary is "periodic" and are usually taken as "stochastic matrices."

3. Restoration of Original Intensity.

If the spatial coefficients and the probability structure of the process are known, for the observation Y , restoration of the original intensity is straightforward using the Bayesian criterion.

Let $X \sim N(\mu, \Gamma_\eta)$ and $\varepsilon \sim N(0, \Gamma_\varepsilon)$ where the block-diagonal matrix $\Gamma_\eta = \{\Sigma_i\}$ and $\Gamma_\varepsilon = \{\Sigma_\varepsilon\}$ are the covariance matrices of the corresponding noise processes. The $\Gamma_\eta = \{\Sigma_i\}$ has a signal-dependent structure, while the blocks of

$\Gamma_\varepsilon = \{\Sigma_\varepsilon\}$ are identical. The likelihood equation for this structure is given by:

$$(\Gamma_\eta^{-1} + \Gamma_\varepsilon^{-1} S' S) X - \Gamma_\varepsilon^{-1} S' A Y - \Gamma_\eta^{-1} \mu = 0$$

where $S = AB$. If $\Phi_s = \Gamma_\eta \Gamma_\varepsilon^{-1} S' S$ and $\gamma(\Phi_s) < 1$ ($\gamma(\cdot)$ denotes the spectral radius of \cdot), then, (Pullman, 1976)

$$X = \sum_{k=0}^{\infty} (-1)^k \Phi_s^k (\Phi_s B^{-1} Y + \mu). \quad (1)$$

In practice, it can be usually presumed that the signal-dependent noise process is quite stable compared to the spatial noise process of the whole image. In addition, the spatial structure of images is generally approximated by coefficients which are center-dominant, and pixels are correlated with others in close spatial proximity. For this reason, a suitable choice of S would typically results in rapid convergence of the terms in (1), so X can be quite adequately estimated by truncating the summation at small k . However, when the coefficients are adaptively estimated, they sometimes turn out to be too irregular to satisfy the requirements of the spectral radius being within unit circle or of fast convergence. It may be useful to apply an iterative approach similar to the point-Jacobian iteration method. Denote $S' S = \{S_{ij}^*\}$, and let $M_0 = \{M_{ij}^0 = \Sigma_i^{-1} + \Sigma_i^{-1} S_{ii}^*\}$ be block diagonal and let $M_1 = \{M_{ij}^1 = \Sigma_i^{-1} S_{ij}^*, i \neq j\}$ be a block matrix of which all diagonal blocks are 0. With the convergence condition $\gamma(M_0^{-1} M_1) < 1$, the original intensities are iteratively recovered: for the ℓ th iteration,

$$X_i^\ell = \Psi_i - (M_i^0)^{-1} \sum_j M_{ij}^1 X_j^{\ell-1} \quad (2)$$

where $\Psi = M_0^{-1} (S' \Gamma_\varepsilon^{-1} A Y) + \Gamma_\eta^{-1} \mu$ is the column vector with n blocks of $\{\Psi_i\}$ which are vectors of order d . The condition to obtain a unique solution are more relaxed than for (1).

4. Estimation of Mean Intensity and Spatial Parameters.

In the proposed system, the parameters used in the restoration of the original intensity are computed adaptively from intensity values predicted on the basis of the previous history of the process.

In remote sensing applications, it is presumed that a "surface patch" in a true scene is likely to be spatially continuous and cohesive. In digital image analysis, the knowledge that pixels which are spatially close together tend to have the same class in which the mean intensities are densely distributed is extremely important. The Bayesian approach is employed herein to estimate the mean intensity as in the restoration of the original intensity. The GRF is used as a measure to quantify the spatial continuity probabilistically, that is, to provide some type of prior information of the mean intensity process.

For a probability structure of μ , a class of Gibbs measures is specified with the energy function $E_p(\mu)$. If $|\cdot|^2$ denotes the square of each of the elements,

$$E_p(\mu) = \sum_{(p,q) \in C_p} \alpha_{p,q} |\mu_p - \mu_q|^2 \quad (3)$$

where $\alpha_{p,q}$ is a nonnegative coefficient vector which represents the "bonding strength" between the p th and q th pixels, and C_p denotes the collection of all pair cliques. A simple homogeneous GRF is assumed in this paper. If the bonding strength is proportional to the proximity of pixel pair in cliques and

is constant across the spectral bands, the energy function is then expressed with scalar α :

$$E_p(\mu) = \alpha \mu' M_p \mu \quad \text{where } M_p = I - D_p$$

where D_p is a proximity matrix consistent with the neighborhood system. The proximity matrix is again a stochastic matrix with all diagonal elements being zero; for the periodic boundary, this matrix is block-circulant. If $X \sim N(\mu, \Gamma_\eta)$, the MAP estimation problem is reformulated as an optimization problem (Lee, 1989):

$$\min (\mathbf{X} - \mu)' \Gamma_\eta^{-1} (\mathbf{X} - \mu) + \alpha \mu' M_p \mu$$

Based on the requirement that the estimate μ must fit the data, the optimal solution of μ can be approximated by using a matrix series expansion as in (1) if a value of some parameter of the distribution of the scaled distance is given. In order to apply this approximation to image processing problems, enough spatial variation in the characteristics of targets associated with the mean intensity should be existed so that $\gamma(\alpha \Gamma_\eta)$ is sufficiently less than 1. Alternatively, the mean intensity can iteratively be estimated with a supervised value of α (Lee, 1989).

The spatial coefficients B and ρ are usually considered to be stationary over time in conventional approaches. This stationarity assumption is not consistent with many remotely sensed processes where temporal changes in the atmospheric environment significantly affect the spatial component.

Let Y be the observation process at the present time. It is complicated to estimate the coefficient T for a model which involves YDY and $YDBX$. This model has a rather complex autoregressive structure. The autoregressive process can be simplified somewhat by taking $S = \{S_{ij}\}$ as an independent parameter. Under this assumption, the interaction among $\{Y_i\}$ is represented by a direct dependence of Y_i on $\{Y_j, j \neq i\}$ according to D and T , and S integrates the spatial relationships of elements of X to explain Y . The spatial component is independent across spectral bands, so the process can be treated separately in each band.

For notational convenience, the band index is omitted in the following discussion. Given X , the log likelihood of a spectral band is given by:

$$n + \log |A| - \frac{n}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} (AY - SX)' (AY - SX) \quad (4)$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Using the properties of the circulant matrix, $SX = M_x S_c$, where the vector $S_c = \{S_0, \dots, S_{n-1}\}$ is a discrete form of the "shift-invariant point spread function" and the matrix M_x is a circular form for arranging X according to S . The dimension of S_c and M_x can be reduced according to the blurring neighborhood system, the order of which is usually chosen to be low. The maximum likelihood estimates of S_c and σ_ε^2 obtained from (4) are

$$\hat{S}_c = (M_x' M_x)^{-1} M_x' A Y \quad \text{and} \quad \hat{\sigma}_\varepsilon^2 = n^{-1} (Y' A' \Phi_x A Y) \quad (5)$$

where $\Phi_x = I - M_x (M_x' M_x)^{-1} M_x'$. By replacing these estimates in (4), the maximum likelihood estimate of ρ then maximizes

$$\frac{2}{n} \log |A| - \log (\Psi_x Y - \rho \Psi_x D Y)' (\Psi_x Y - \rho \Psi_x D Y).$$

The likelihood is evaluated to choose a value of ρ within the bounds ± 1 . Denoting eigenvalues of the stochastic matrix D as $\{\lambda_i\}$, the first term is expanded

$$\frac{2}{n} \log |A| = -\frac{2}{n} \rho \sum_i \lambda_i - \frac{1}{n} \rho^2 \sum_i \lambda_i^2 - \frac{2}{n} \sum_{k=3}^n \frac{\rho^k}{k} \sum_i \lambda_i^k.$$

These eigenvalues are easily computed by the fast Fourier transform and $|\lambda_i| \leq 1$. If the function is searched in the region $|\rho| < 1$, the terms which are functions of at least the third power of the eigenvalues are trivial. Then, the optimal value is equivalent to the solution of the equation

$$\rho^3 + (\phi_0 - 2\psi_1)\rho^2 + (\phi_1 - 2\phi_0\psi_1 + \psi_0)\rho + (\phi_0\psi_0 - \phi_1\psi_1) = 0$$

where

$$\phi_0 = \sum_i \lambda_i / \sum_i \lambda_i^2, \quad \phi_1 = n / \sum_i \lambda_i^2, \\ \psi_0 = \frac{Y' \Psi_x Y}{Y' D' \Psi_x D Y}, \quad \psi_1 = \frac{Y' \Psi_x D Y}{Y' D' \Psi_x D Y}.$$

A simple line search can find a zero point corresponding the maximum likelihood within the bounds of ρ .

5. Adaptive Prediction.

The conventional matrix-oriented algorithms for adaptive linear prediction of multivariate processes are susceptible to numerical instability due to matrix inversions and cause "run time and storage complexity" for multiple line processes such as pixel-by-pixel image processing. An alternative approach has been implemented for the image reconstruction system of multitemporal images. The method based on the "orthogonal escalator structure" (Ahmed and Youn, 1980) requires only scalar operations and less storage for the parameter updating procedures.

Let $Z_t = \{z_t^i\}$ be a stationary zero-mean multivariate time series of order d . If the series is a linear autoregressive moving-average (ARMA) process of order (p, q) ,

$$Z_t = C V_t + \eta_t$$

$$C = \{C_1^z, \dots, C_p^z, C_1^\eta, \dots, C_q^\eta\}$$

$$V_t = \{Z_{t-1}^z, \dots, Z_{t-p}^z, \eta_{t-1}^\eta, \dots, \eta_{t-q}^\eta\}.$$

$C = \{c_{ij}\}$ is the $d \times K$ matrix and $V_t = \{v_t^i\}$ is the vector of order K where $K = d(p+q)$. A linear predictor of ARMA (p, q) based on V_t has the form

$$\hat{Z}_t = \hat{C} V_t.$$

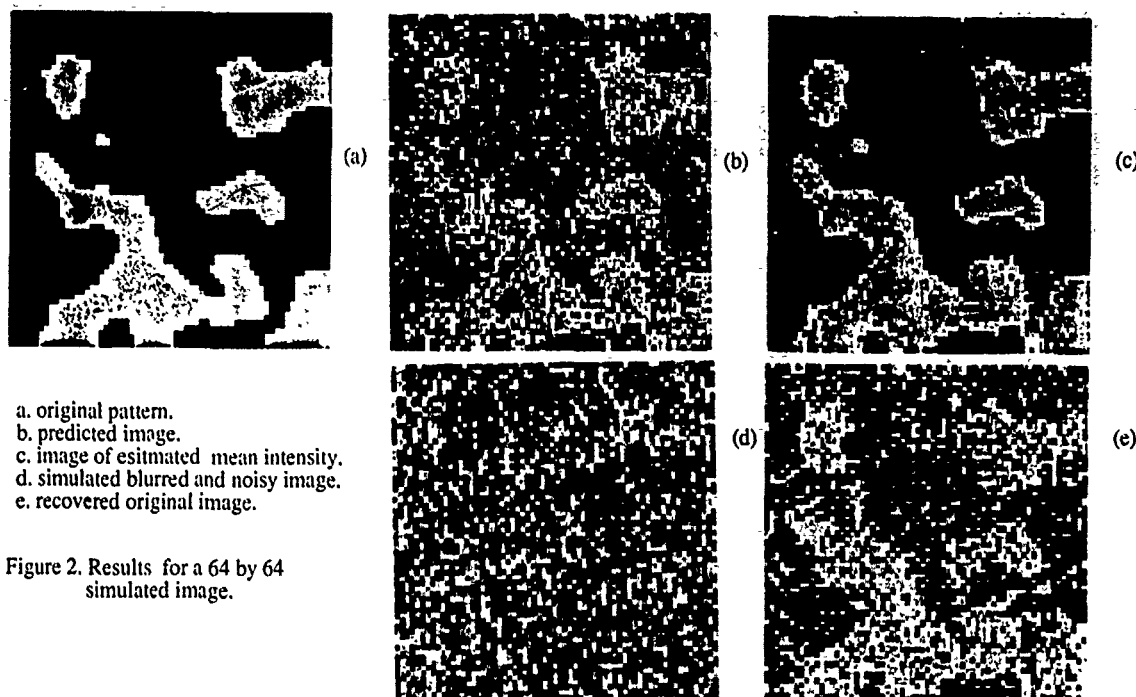
The adaptive coefficients are expressed as a function of time due to their time-varying characteristics: $C_t = \{c_{ij}^t\}$.

Given the covariance matrix Σ_v^t of V_t , there exists a unit lower triangular matrix L such that

$$\Lambda_v^t = L \Sigma_v^t L'.$$

The diagonal elements of Λ_v^t are the characteristic roots of the covariance matrix and the elements of the vector $W_t = L V_t = \{w_t^i\}$ are uncorrelated. The key feature of the algorithm is a factorization procedure related to the unit lower triangle transform. The matrix L can be decomposed into a product of unit lower triangular matrices

$$L = \prod_{k=1}^K L_k$$



a. original pattern.
b. predicted image.
c. image of estimated mean intensity.
d. simulated blurred and noisy image.
e. recovered original image.

Figure 2. Results for a 64 by 64 simulated image.

$L_t = \{l_{ij}^k | l_{ii}^k = 1, l_{ij}^k = l_{ji}^k \text{ and all another elements are } 0\}$.

Using the triangular structure, W_t is obtained recursively

$$w_{ij}^t = \begin{cases} r_{ij}^t, & j = 1 \\ w_{i,j-1}^t - l_{i,j-1}^t w_{j-1}^t, & j \neq 1 \text{ and } i \geq j \\ w_i^t = w_{ii}^t, & i = 1, \dots, K. \end{cases}$$

Using the adaptive least-squares escalator predictor developed by Ahmed and Youn (1980), l_{ij}^k can be updated. The adaptive least-squares linear predictor for the actual process Z_t then results from the alternative representation of the forecasting equation

$$\hat{Z}_t = \hat{C}_t L^{-1} W_t = \hat{C}_t W_t.$$

The parameters are updated using a steepest descent approach (Hong and Messerschmitt, 1984) according to

$$\begin{aligned} \hat{c}_{ij}^{t+1} &= \hat{c}_{ij}^t + \pi_c^t \eta_i^t w_j^t \\ \lambda_j^{t+1} &= \zeta \lambda_j^t + (1 - \zeta) w_j^{t^2} \end{aligned}$$

where $\pi_c^t = \pi / \lambda_j^{t+1}$, $\eta_i^t = z_i^t - \sum_{j=1}^K \hat{c}_{ij}^t w_j^t$. The convergence rate π and the smoothing parameter ζ are bounded between 0 and 1, and generally chosen close to 1. The covariance matrix of \hat{Z}_t is

$$E[\hat{Z}_t \hat{Z}_t^T] = \hat{C}_t \Lambda_t^t \hat{C}_t^T.$$

If the recent data come from a different distribution, the likelihood variable

$$q_t = \sum_{j=1}^K [\log \lambda_j^t + w_j^{t^2} / \lambda_j^t]$$

will increase significantly so q_t can be used as a good detection statistic for the "unexpectedness" of the recent data points in the variance sense.

The n -pixel intensity process after decorrelating the spatial components is considered as n independent time processes where Z_t corresponds to $X_{(t),i} - \mu_{(t),i}$, $i = 1, \dots, n$. The local stationarity of the mean intensity can usually be assumed

in multitemporal processes over short observation intervals. Otherwise, its variation can be adapted by fitting to a simple polynomial function of time with "exponential weight" because of the independence of the mean intensity (Lee, 1989).

6. Examples.

The new adaptive reconstruction system has been tested with simulation data. Figure 2 illustrates examples of the results for a simple simulation image. The data were generated from autoregressive stationary time series of second order for 30 time steps and spatially contaminated according to the model given in Section 2. The original pattern of the whole series was modelled as a GRF of 3 classes. A full analysis for the results of the algorithm when applied to simulation data and to satellite images will be presented a forthcoming paper.

Reference.

- Ahmed, N. and Youn, D. H., "On a realization and related algorithm for adaptive prediction," *IEEE Trans. Acoust., Speech, Signal Proc.* Vol ASSP-28, 493-497, 1980.
- Georgii, H. O., *Canonical Gibbs Measure*. Springer-Verlag, Berlin, 1979.
- Hong, M. L. and Messerschmitt, *Adaptive Filters*. Kluwer Academic Publishers, Hingham, 1984.
- Khazenic, N., "Contextual classification of remotely sensed data, based on a spatial temporal correlation model," *Ph. D. Dissertation*. Univ. of Texas at Austin, 1987.
- Lee, S., "Statistically based unsupervised classification and adaptive estimation in multispectral, multitemporal image processing," *Ph. D. Dissertation*. Univ. of Texas at Austin, 1989.
- Pullman, N. J., *Matrix Theory and Its Applications*. Marcel Dekker Inc., New York, 1976.
- Ripley, B. D., *Spatial Statistics*. Wiley, New York, 1981.
- Robinson, E. A., *Multichannel Time Series Analysis with Digital Computer Programs*. Holden-Day, San Francisco, 1967.

A METHOD OF STATISTICAL MULTISOURCE CLASSIFICATION WITH A MECHANISM TO WEIGHT THE INFLUENCE OF THE DATA SOURCES

J.A. Benediktsson and P.H. Swain

School of Electrical Engineering

and

Laboratory for Applications of Remote Sensing

Purdue University

W. Lafayette, IN 47907, U.S.A.

Telephone: (317) 494 - 1743, FAX: (317) 494 - 6440

ABSTRACT

Statistical multisource classification of remote sensing data and geographic information by means of a method based on Bayesian classification theory is investigated. Extensions are made to the method to control the influence of the data sources involved in the classification process. Reliability measures are introduced to rank the quality of the data sources. The data sources are then weighted according to these rankings in the multisource classification. Four data sources are used in experiments: One is Landsat MSS data; the other three contain topographic data (elevation data, slope data and aspect data). The data classes in the Landsat MSS data are treated as Gaussian but the data classes in the other sources are treated as non-Gaussian.

KEY WORDS: Classification, multisource data, global membership function, reliability, weights.

1. INTRODUCTION

Computerized information extraction from remotely sensed imagery has been applied successfully over the last two decades. The image data has mostly been multispectral data and the statistical pattern recognition (multivariate classification) methods are now widely known. Within the last decade advances in space and computer technologies have made it possible to amass large amounts of data about the Earth and its environment. The data are now more and more typically not only spectral data but include, for example, forest maps, ground cover maps, radar data and topographic information such as elevation and slope data. We may therefore have many kinds of data from different sources regarding the same scene. These are collectively called multisource data.

We are interested in using all these data to extract more information and get more accuracy in classification. However, the conventional multivariate classification methods cannot be used satisfactorily in processing multisource data. This is due to several reasons. One is that the multisource data cannot be modeled by a convenient multivariate statistical model since the data are multitype. They can for example be spectral data, elevation ranges and even non-numerical data such as ground cover classes or soil types. The data are also not necessarily in common units and therefore scaling problems may arise. It is also desirable to determine the reliability of each data source and weight the sources in the classification, because all the sources are in general not equally reliable. This all implies that methods other than the conventional multivariate classification have to be used to classify multisource data.

Various ad hoc methods have been proposed to classify multisource data. However, we are interested in developing more general methods which can be applied to classify any type of data. In particular, our attention is focused on statistical multisource analysis by means of an approach based on Bayesian classification theory which was proposed by Swain, Richards and Lee [1,2]. An extension of their

method will be examined in this paper.

Our objective is to modify the method to take into account the relative reliabilities of the sources of data involved in the classification. This requires a way to quantify the reliability of a data source, which becomes important when we look at the combination of information. The foundation of the method for combination from various sources consists essentially of multiplication of source-specific posterior probabilities from all the sources involved in the classification. If any of the sources are unreliable they can affect the outcome of the multiplication disproportionately and consequently increase classification error.

In this paper we investigate methods to determine the reliability and define a corresponding reliability factor (weight) for each data source. The reliability factors are then included in the classification process. Experimental results are given.

2. FUNDAMENTALS OF STATISTICAL MULTISOURCE ANALYSIS

The method proposed in [1,2] extends well-known concepts used for classification of multispectral images involving only one data source. In this method the various data sources are handled independently and each can be characterized by any appropriate model. The data in each source is first classified into a source-specific number of data classes. The information from the sources is then aggregated by a global membership function and the data is classified according to the usual maximum selection rule into a user-specified number of information classes. The global membership function has the following form for the information class ω_j if n data sources are used:

$$F_j(X) = [p(\omega_j)]^{1-n} \prod_{s=1}^n p(\omega_j | x_s) \quad (1)$$

where $X = [x_1, x_2, \dots, x_n]$ is a pixel, $p(\omega_j)$ is the prior probability of ω_j and $p(\omega_j | x_s)$ is a source-specific posterior probability.

This statistical multisource analysis approach is an extension of single-source Bayesian classification. However, the method as presented by Swain, Richards and Lee [1,2] does not provide a satisfactory mechanism to account for varying degrees of reliability. It is our belief that this problem can be overcome if reliability factors are associated with each source involved in the classification. For this reason we will investigate a modified version of this method by means of which reliability analysis is added to the classification process.

3. THE RELIABILITY APPROACH

We want to associate reliability factors with the sources in the global membership function discussed above, i.e., to express quantitatively our confidence in each source, and use the reliability factor for classification purposes. This is very important because it is necessary to increase the influence of the "more reliable" sources, i.e., the sources we have more confidence in, on the global membership function and consequently decrease the influence of the "less reliable" sources in order to improve the classification accuracy. The need for reliability factors becomes apparent by looking at equation (1) where the global

* This research is supported by the National Aeronautics and Space Administration Contract No. NAGW-925

membership function is a product of probabilities related to each source. Each probability has value in the interval from 0 to 1. If any one of them is near zero it will carry the value of the membership function close to zero and therefore downgrade drastically the contribution of information from other sources, even though the particular source involved may have little or no reliability.

It is clear then that it is desirable to put weights (reliability factors) on the sources which will influence their contributions to classification. Since the global membership function is a product of probabilities this weight has to be involved in such a way that when the reliability of a source is low it must discount the influence of that source and when the reliability of a source is high it must give the source relatively high influence. One possible choice for this kind of analysis is to use reliability factors as exponents on the factor for each source in the global membership function.

Let us now determine the contribution from a single source in the global membership function. The global membership function for n sources is shown in equation (1). If one source is added to those n sources, i.e., we have $n + 1$ sources, the global membership function could be written in the following form:

$$F_j(X) = [p(\omega_j)]^{-n+1} \prod_{i=1}^{n+1} p(\omega_j | x_i) \quad (2)$$

If equation (2) is divided by equation (1) we get the contribution from source number $n+1$ which is $p(\omega_j | x_{n+1})/p(\omega_j)$. This motivates us to rewrite equation (1) in the following form:

$$F_j(X) = p(\omega_j) \prod_{i=1}^n \{p(\omega_j | x_i)/p(\omega_j)\} \quad (3)$$

Now to control the influence of each source, reliability factors α_i are assigned as exponents on the contribution from each source. Therefore equation (3) with reliability factors is written as:

$$F_j(X) = p(\omega_j) \prod_{i=1}^n \{p(\omega_j | x_i)/p(\omega_j)\}^{\alpha_i} \quad (4a)$$

where the α_i 's ($i = 1, \dots, n$) are selected in the interval $[0,1]$ because of the following reasons. If source i is totally unreliable ($\alpha_i = 0$) it will not have any influence on (4a) because

$$\{p(\omega_j | x_i)/p(\omega_j)\}^0 = 1$$

regardless of the value of $p(\omega_j | x_i)$. And if source i has the highest reliability ($\alpha_i = 1$) then it will give a full contribution to (4a) because

$$\{p(\omega_j | x_i)/p(\omega_j)\}^1 = p(\omega_j | x_i)/p(\omega_j)$$

It is also worthwhile to note that this method of putting exponents on the probabilities does not change the decision for a single-source classification because the exponential function p^α is a monotonic function of p .

Equation (4a) can also be written in a logarithmic form as:

$$\log F_j(X) = \log p(\omega_j) + \sum_{i=1}^n \alpha_i \log \{p(\omega_j | x_i)/p(\omega_j)\} \quad (4b)$$

where the reliability factors are expressed as the coefficients in the sum. These coefficients act like weights in the sum and control the influence of each source on the global membership function. Another way to see this is to look at the sensitivity of the global membership function to changes in one of the probability ratios. This can be expressed as:

$$\frac{\delta F_j(X)}{F_j(X)} = \alpha_i \frac{\delta p(\omega_j | x_i)/p(\omega_j)}{p(\omega_j | x_i)/p(\omega_j)} \quad (5)$$

and shows that α_i represents the sensitivity which implies that the value of α_i will control the influence of source number i on the global membership function since a percentage change in the posterior probability leads to the same percentage change in the global membership function, multiplied by α_i .

The problem is to quantify the reliability of the sources and define the reliability factors, $\{\alpha_i\}$, based on the reliability of the sources. We think of a source as being reliable if its contribution to the combination of information from various sources is "good", i.e., if the classification accuracy is increased substantially or more information is extracted by using this particular source.

The process of determining the reliability factors is a two stage process. First the reliabilities of the sources have to be measured by some appropriate "reliability measure" and then the values of the reliability measures must be associated with the reliability factors in the global membership function.

3.1 Reliability Measures

Using the above understanding of a reliable source, three measures are applied to quantify the reliability of a source: weighted average separability, overall classification accuracy and equivocation. All these measures are related to the classification accuracy of the source.

3.1.1 Separability of Information Classes

We call a source reliable if the separability of the information classes is high for the source. If on the other hand the separability of the information classes is low, the source is not reliable. Therefore one possibility for reliability evaluation is to use the average statistical separability of the information classes in each source, e.g., average Bhattacharyya distance, average Jeffries-Matusita (JM) distance, average transformed divergence or any other separability function [3]. What kind of average is used depends on what we are after in the multisource classification. For instance, if it is desired to improve the overall classification accuracy the overall average is used. If, however, we are concentrating on specific classes, a weighted average separability of those information classes is used. Using separability with the Gaussian assumption has the disadvantage that then it is usually necessary to compute covariance matrices to estimate the separability. Covariance matrices do not always exist, e.g., for some topographic classes. Also, if the Gaussian assumption is not made it becomes more difficult to compute the separability.

3.1.2 Classification Accuracy of a Data Source

Another way to measure reliability of a data source is to use the classification accuracy of the source. In this case a source is considered reliable if the classification accuracy for the source is high, but if the accuracy is low the source is considered unreliable. This approach is related to the method of using separability measures in that increased separability is consistent with higher accuracy. On the other hand there is no need estimate covariance matrices to compute the classification accuracy so this approach is always applicable.

3.1.3 Equivocation

Still another way to characterize reliability of a source is to examine how strongly the data classes indicate information classes, i.e., by looking at the conditional probabilities that a specific information class is observed given a data class. All these conditional probabilities can be computed by comparing the reference map to a classification map from a data source.

Assuming there are M information classes $\{\omega_1, \dots, \omega_M\}$ and m data classes $\{d_1, \dots, d_m\}$, the conditional probabilities can be written as the $m \times M$ correspondence matrix R , where R is:

$$R = \begin{bmatrix} p(\omega_1 | d_1) & p(\omega_2 | d_1) & \dots & p(\omega_M | d_1) \\ p(\omega_1 | d_2) & p(\omega_2 | d_2) & \dots & p(\omega_M | d_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(\omega_1 | d_m) & p(\omega_2 | d_m) & \dots & p(\omega_M | d_m) \end{bmatrix} \quad (6)$$

Reliability can now be defined in the following way: If a source is optimal in reliability there would be a unique information class corresponding to each data class. Therefore ideally one conditional probability in each row of R would be 1 and all the others would be zero. If a source were very unreliable, there would be no correspondence between the data classes and the information classes; in the worst case all the numbers in the matrix would be equal.

Now it is necessary to associate a number with the matrix R to characterize the reliability. Using information theoretic measures [4] the information classes can be thought of as transmitted signals and the data classes as received signals which must be used to estimate the transmitted signals. Using this approach it can be stated that there is an uncertainty of $\log[1/p(\omega_i | d_j)]$ about the information class ω_i when data class d_j is observed in a data source.

The average loss of information can be calculated when the data class d_j is observed, which is given by:

$$H(\omega|d_j) = \sum_i p(\omega_i|d_j) \log \frac{1}{p(\omega_i|d_j)} \quad (7)$$

Now we want to average the information loss over all observed data classes d_j . This is the equivocation of ω with respect to d and is denoted by $H(\omega|d)$:

$$\begin{aligned} H(\omega|d) &= \sum_j p(d_j) H(\omega|d_j) \\ &= \sum_j \sum_i p(d_j) p(\omega_i|d_j) \left\{ \log \frac{1}{p(\omega_i|d_j)} \right\} \\ &= \sum_i \sum_j p(\omega_i, d_j) \left\{ \log \frac{1}{p(\omega_i|d_j)} \right\} \end{aligned} \quad (8)$$

$H(\omega|d)$ represents the average uncertainty about an information class over all the data classes. Evidently, $H(\omega|d)$ is the average loss of information per data class and therefore seems to be a reasonable term to associate with the reliability of a source. Since $H(\omega|d)$ measures uncertainty, the lower value it has the more reliable a source is. Therefore, the equivocation is called an uncertainty measure rather than a reliability measure. To be able to transform this uncertainty measure into a reliability factor it first has to be mapped into a reliability measure and then associated with a reliability factor.

3.2 Association

The values of the reliability (uncertainty) measures must be associated with the reliability factors in order to improve the classification accuracy. It is worthwhile to note that we only want to include sources in the global membership function if the presence of those sources improves the classification accuracy, i.e., we want the classification accuracy to be an increasing function of the number of sources. This is similar to feature selection but the difference here is that the sources (features) are not only selected but also the contribution of each source to the global membership function is quantified.

Any of the measures discussed in section 3.1 gives a specific value which should be mapped into a reliability factor on the basis of the strength of our belief in the contribution of the source to the classification accuracy. The reliability (or uncertainty) measures take values in some particular interval and it is necessary to know the (functional) mapping between the values of the measures and the values of the reliability factors. It is very difficult to find an explicit association function between the values of the reliability and uncertainty measures on one hand and the reliability factors on the other. The measures can easily be used to rank the sources from "best" to "worst" but it is difficult to determine the optimal value of the reliability factors. Ranking measures have previously been used in *consensus theory* [5] but that has mostly been for linear opinion pool problems whereas in contrast the global membership function (4a) can be considered an *independent opinion pool* problem.

In this paper the reliability and uncertainty measures will be used to rank the sources. Several different reliability factors will be used in the experiments and we will investigate the relationships between the values of the reliability factors and the overall classification accuracy after combination of the sources.

4. EXPERIMENTAL RESULTS

The statistical multisource classification method was used to classify a data set consisting of the following 4 data sources:

- 1) Landsat MSS data (4 data channels)
- 2) Elevation data (in 10 m contour intervals, 1 data channel)
- 3) Slope data (0-90 degrees in 1 degree increments, 1 data channel)
- 4) Aspect data (1-180 degrees in 1 degree increments, 1 data channel)

Each channel is 135 rows and 131 columns, all co-registered.

The area used for classification is a mountainous area in Colorado. It has 10 ground cover classes which are listed in table 1. One class is water; the others are forest types. It is very difficult to distinguish between the forest types using the Landsat MSS data alone since the forest classes show very similar spectral response. With the help of elevation, slope and aspect they can better be distinguished.

Ground reference data were gathered for the area by comparing a cartographic map to a color composite of the Landsat data and also comparing it to a line printer output of each Landsat channel. By this method 2019 ground reference points (11.4% of the area) were selected. Ground reference consisted of two or more homogenous fields (regions) in the imagery for each class. For each class the largest region was selected as a training region and the other regions were used for testing the classifiers. Overall 1188 pixels were used for training and 831 for testing the classifiers.

To satisfy the underlying assumptions of the statistical multisource algorithm we need to show that the data sources can be treated independently in the classification. We do that by looking at the correlations between all seven data channels in table 2. The correlations between the sources are in all cases low, so we can treat the data sources as independent and use the global membership function in (4a) and (4b) as the classifier.

Each source was treated independently in training. The data classes in the Landsat MSS source were modeled by the Gaussian distribution, where the mean vectors and covariance matrices were estimated from the training fields. The other data sources have non-Gaussian data classes. There the histograms of the training fields were used to estimate the density functions for these sources.

Statistical multisource classification was performed on the data with varying weights (reliability factors) for the data sources. The results of classification of the training fields are shown in table 3 and for the test fields in table 4. The reliability and uncertainty measures introduced in section 3.1 were used to rank the data sources. Looking at the classification results for the specific data sources (in both tables 3 and 4), we see that the Landsat data is the most reliable source, elevation second, aspect third and slope the least reliable. This is the same ranking indicated by the equivocation measure in table 5. (We cannot use the separability measures with the Gaussian assumption here, since some of the data classes in the topographic sources have singular covariance matrices.) In all the experiments the Landsat data were given the highest weights while the weights of the other sources were varied.

If we look at the classification of the training samples (table 3) we see that by combining all the sources with equal weights we improve the overall classification accuracy to 74.2%, i.e., by more than 6% compared to the best accuracy in the single source classifications (Landsat MSS: 67.9%). By lowering the weights on the topographic sources to a certain point we get the overall accuracy up to 78.0%. Therefore by changing the weights of the sources we can improve the classification accuracy of the training samples by 3.8%. This "best" result is achieved when the Landsat source has full weight and the other sources each have 40% weight. It is also nearly achieved when the Landsat source has full weight, the elevation source has 50% weight, the aspect source has 40% weight and the slope source has 30% weight, which is a weighting suggested by the ranking of the reliability measures. Using some other weights suggested by the reliability measure also gives a very good result (77.8% overall accuracy). In general the results in table 3 show that we can improve the overall classification accuracy by discounting the influence of some of the data sources. In table 3 we also see that if we discount the data sources too much, the overall classification accuracy goes down, as could be expected.

If we look at the results in table 4 we see very similar results to the ones in table 3. The results in table 4 show the results of the classification of test fields and therefore the classification accuracy is lower than in table 3. If the sources all have equal weights, then the overall accuracy is 56.0% which is up 2.9% from the classification accuracy of the best single source (Landsat MSS: 53.1%). This is not as much increase as in the case of classification of training data. Using different levels of reliability we improve the overall classification accuracy to 60.8% which is 4.8% more than with the equal weight approach. We get this best result when the Landsat source has full weight, the elevation source 80% weight, the aspect source 60% weight and the slope source 40% weight. This particular weighting is suggested by the ranking measures. 60.8% overall accuracy is very good for the classification of test fields in this area.

5. CONCLUSIONS

The statistical multisource analysis method proposed in [1,2] and modified in this paper worked well for combining multispectral and

topographic data in our experiments. The combination of four data sources gave significant improvement in overall classification accuracy compared to single source classification. Using different levels of weights for different sources also showed promise in our experiments in terms of increase in overall classification accuracy. By assigning specific weights (reliability factors) to the data sources, overall classification accuracy was improved about 4% to 5% from the overall classification accuracy for which equal weights were used.

Two of the suggested reliability measures were employed as ranking criteria for the data sources. These worked well but the problem remains: how to find the optimal weights for the data sources. A goal of current research is to frame this problem as an optimisation problem and solve it by linear programming using the training samples. We are also investigating different methods for modeling non-Gaussian data since the histogram approach applied in this paper is not considered sufficient.

REFERENCES

1. P.H. Swain, J.A. Richards and T. Lee, "Multisource Data Analysis in Remote Sensing and Geographic Information Processing," *Proceedings of the 11th International Symposium on Machine Processing of Remotely Sensed Data 1985*, West Lafayette, Indiana, pp. 211-217, June 1985.
2. T. Lee, J.A. Richards and P.H. Swain, "Probabilistic and Evidential Approaches for Multisource Data Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, no. 3, pp. 283-293, May 1987.
3. J.A. Richards, *Remote Sensing Digital Image Analysis - An Introduction*, Springer-Verlag, Berlin, W. Germany 1986.
4. C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Chicago 1963.
5. R.L. Winkler, "The Consensus of Subjective Probability Distributions", *Management Science*, vol. 15, no. 2, Oct. 1968, pp. B-61 - B-75.

TABLE 1

| Class # | Information Class |
|---------|----------------------------------|
| 1 | water |
| 2 | colorado blue spruce |
| 3 | mountane/subalpine meadow |
| 4 | aspen |
| 5 | ponderosa pine |
| 6 | ponderosa pine/douglas fir |
| 7 | engelmann spruce |
| 8 | douglas fir/white fir |
| 9 | douglas fir/ponderosa pine/aspen |
| 10 | douglas fir/white fir/aspen |

TABLE 2

Correlations between the data channels.
(1,2,3 and 4 are Landsat MSS channels 1 to 4,
5 is elevation, 6 is slope and 7 is aspect.)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|------|------|------|-------|------|-------|-------|
| 1 | 1.00 | 0.68 | 0.53 | 0.15 | 0.12 | 0.01 | -0.08 |
| 2 | | 1.00 | 0.50 | 0.20 | 0.13 | 0.02 | -0.08 |
| 3 | | | 1.00 | -0.04 | 0.05 | -0.05 | -0.02 |
| 4 | | | | 1.00 | 0.23 | 0.15 | -0.17 |
| 5 | | | | | 1.00 | 0.17 | -0.19 |
| 6 | | | | | | 1.00 | -0.00 |
| 7 | | | | | | | 1.00 |

TABLE 3

Classification results of training samples for four data sources and composites with various values of the reliability factor in statistical multisource analysis. Here we combine the Landsat MSS, elevation, slope and aspect data.

| * l e s a | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|--------------|--|-----|----|----|-----|-----|-----|----|----|----|------|
| Landsat MSS | 99 | 48 | 0 | 80 | 9 | 69 | 92 | 0 | 0 | 0 | 67.9 |
| elevation | 100 | 0 | 0 | 23 | 17 | 13 | 98 | 0 | 16 | 20 | 58.4 |
| slope | 100 | 0 | 0 | 0 | 5 | 64 | 0 | 0 | 0 | 0 | 41.5 |
| aspect | 100 | 0 | 0 | 44 | 12 | 15 | 59 | 0 | 0 | 0 | 53.6 |
| 1. 1. 1. 1. | 100 | 98 | 0 | 35 | 35 | 80 | 100 | 0 | 0 | 0 | 74.2 |
| 1. .5 .5 .5 | 100 | 99 | 0 | 65 | 34 | 76 | 94 | 0 | 0 | 62 | 77.6 |
| 1. .4 .4 .4 | 100 | 100 | 11 | 71 | 33 | 73 | 95 | 0 | 0 | 58 | 78.0 |
| 1. .3 .3 .3 | 100 | 100 | 11 | 75 | 27 | 71 | 96 | 0 | 0 | 42 | 76.9 |
| 1. .2 .2 .2 | 100 | 98 | 11 | 75 | 23 | 71 | 96 | 0 | 0 | 26 | 75.5 |
| 1. .1 .1 .1 | 100 | 96 | 18 | 75 | 15 | 66 | 97 | 38 | 0 | 0 | 74.2 |
| 1. .8 .4 .6 | 100 | 99 | 0 | 64 | 37 | 79 | 93 | 0 | 0 | 60 | 77.8 |
| 1. .8 .1 .2 | 100 | 100 | 11 | 74 | 17 | 76 | 95 | 0 | 0 | 35 | 76.0 |
| 1. .6 .4 .5 | 100 | 99 | 4 | 67 | 34 | 76 | 94 | 0 | 0 | 60 | 77.8 |
| 1. .5 .3 .4 | 100 | 100 | 11 | 73 | 33 | 75 | 95 | 0 | 0 | 49 | 77.9 |
| 1. .4 .2 .3 | 100 | 100 | 11 | 75 | 27 | 73 | 96 | 0 | 4 | 38 | 77.0 |
| 1. .3 .1 .2 | 100 | 99 | 11 | 75 | 18 | 74 | 96 | 0 | 4 | 22 | 75.4 |
| # of pixels | 408 | 88 | 45 | 75 | 105 | 126 | 224 | 32 | 25 | 60 | 1188 |

*
l e s a indicates the weights assigned to the Landsat MSS (l), elevation (e), slope (s) and aspect (a) sources.

TABLE 4

Classification results of test samples for four data sources and composites with various values of the reliability factor in statistical multisource analysis. Here we combine the Landsat MSS, elevation, slope and aspect data.

| * l e s a | Percent Agreement with Reference for Class | | | | | | | | | | OA |
|--------------|--|----|----|----|-----|-----|-----|----|----|----|------|
| Landsat MSS | 97 | 0 | 0 | 0 | 25 | 79 | 97 | 0 | 0 | 0 | 53.1 |
| elevation | 100 | 0 | 0 | 20 | 2 | 21 | 100 | 0 | 8 | 21 | 40.4 |
| slope | 86 | 0 | 0 | 0 | 0 | 5 | 33 | 0 | 0 | 0 | 24.3 |
| aspect | 95 | 0 | 0 | 15 | 1 | 6 | 19 | 0 | 0 | 0 | 26.7 |
| 1. 1. 1. 1. | 86 | 0 | 0 | 25 | 35 | 92 | 86 | 0 | 0 | 0 | 56.0 |
| 1. .5 .5 .5 | 86 | 0 | 0 | 48 | 45 | 80 | 97 | 0 | 0 | 0 | 57.9 |
| 1. .4 .4 .4 | 86 | 0 | 0 | 52 | 49 | 76 | 97 | 0 | 0 | 0 | 57.9 |
| 1. .3 .3 .3 | 86 | 0 | 0 | 54 | 51 | 63 | 97 | 0 | 0 | 44 | 57.4 |
| 1. .2 .2 .2 | 97 | 0 | 0 | 0 | 54 | 80 | 97 | 0 | 0 | 31 | 59.5 |
| 1. .1 .1 .1 | 93 | 0 | 0 | 0 | 54 | 76 | 97 | 0 | 0 | 26 | 57.3 |
| 1. .8 .4 .6 | 100 | 0 | 0 | 51 | 38 | 84 | 97 | 0 | 0 | 0 | 60.8 |
| 1. .8 .1 .2 | 91 | 0 | 0 | 60 | 48 | 72 | 97 | 0 | 0 | 0 | 58.6 |
| 1. .6 .4 .5 | 86 | 0 | 0 | 51 | 44 | 81 | 97 | 0 | 0 | 0 | 58.0 |
| 1. .5 .3 .4 | 86 | 0 | 0 | 54 | 48 | 74 | 97 | 0 | 0 | 0 | 57.5 |
| 1. .4 .2 .3 | 97 | 0 | 0 | 57 | 51 | 55 | 97 | 0 | 0 | 41 | 58.2 |
| 1. .3 .1 .2 | 95 | 0 | 0 | 0 | 55 | 80 | 97 | 0 | 0 | 33 | 59.3 |
| # of pixels | 195 | 24 | 42 | 65 | 139 | 188 | 70 | 44 | 25 | 39 | 831 |

*
l e s a indicates the weights assigned to the Landsat MSS (l), elevation (e), slope (s) and aspect (a) sources.

TABLE 5

Equivocation of the data sources

| Source | Equivocation | Rank |
|-----------|--------------|------|
| MSS | 0.216955 | 1 |
| Elevation | 0.252676 | 2 |
| Aspect | 0.277244 | 3 |
| Slope | 0.289636 | 4 |

OBSERVATIONS OF CLOUDS AND FOG WITH A 1.4 MM WAVELENGTH RADAR

James B. Mead and Robert E. McIntosh

University of Massachusetts
Microwave Remote Sensing Laboratory
Amherst, MA 01003

Abstract

A portable 1.4 mm wavelength (215 GHz) incoherent radar having 60 W output power has been developed by the University of Massachusetts Microwave Remote Sensing Laboratory. The short wavelength and high directivity of this radar indicate its usefulness for high spatial resolution atmospheric sensing of low reflectivity targets. The radar has sufficient sensitivity to detect -25 dBZ targets at a range of 1 km. Measurements of cloud and fog reflectivity were made for ranges between 36 and 47 meters which are, to our knowledge, the first such measurements made by a radar operating at this wavelength. Herein we describe the instrument, predict target detectability and give preliminary results of fog and cloud scattering experiments.

System Description

A block diagram of the 1.4 mm radar system is given in Figure 1. Table 1 summarizes the important parameters for estimating overall system performance. The transmitter employs a Varian VKY2429M1 Extended Interaction Oscillator (EIO) ca-

pable of producing 60 W pulses at a duty cycle up to 0.005. The modulator consists of an internal Pulse Repetition Frequency (PRF) generator, a Field Effect Transistor (FET) switch, and a high-voltage triode switch. The triode acts as a hard-tube modulator that is capable of providing the high peak currents to rapidly charge the EIO stray capacitance, while providing a flat pulse to minimize EIO frequency drift. The receiver front-end employs a single-ended mixer driven by a 71.2 GHz InP Gunn diode local oscillator through a frequency tripler. The double sideband (DSB) noise figure of the mixer subsystem is 10 dB. An automatic frequency control (AFC) loop operating between 1.2 and 1.6 GHz down-converts the signal to 160 MHz, where a logarithmic amplifier is used prior to detection. Separate 6-inch Gaussian optics lens antennas with scalar feed horns are used to achieve high isolation (> 100 dB) between transmitter and receiver. A dual antenna scheme was chosen due to inefficiency of circulators and transmit/receive switches at this wavelength.

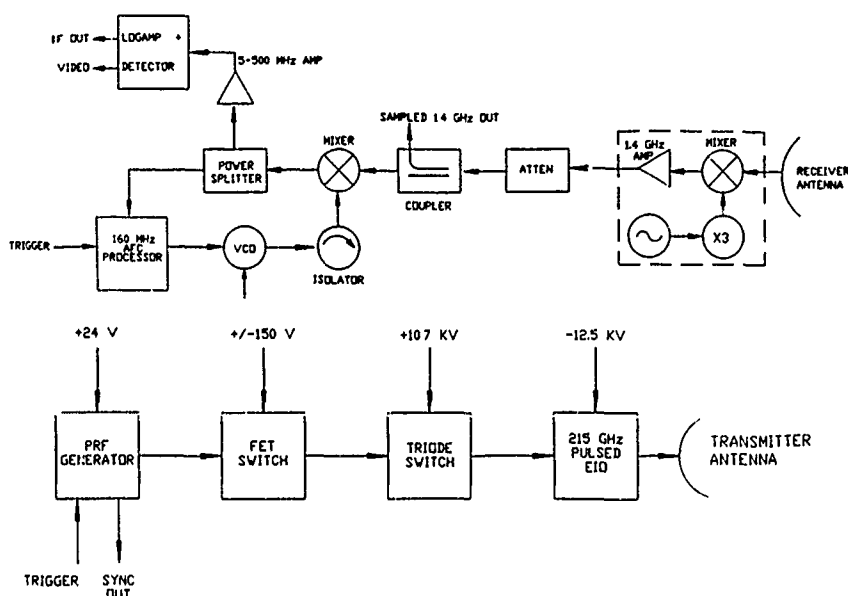


Figure 1. Radar block diagram.

Table 1. Characteristics of UMass 1.4 mm wavelength radar system.

| Transmitter | |
|-------------------|--------------------|
| Center Frequency | 215 GHz Nominal |
| Peak Output Power | 60 W |
| Pulsewidth | 100-500 nsec |
| PRF | 700 Hz-20 KHz |
| Tuning Bandwidth | 300 MHz |
| Max Duty Cycle | .005 |
| Antenna | |
| | 6" Lens |
| 3 dB beamwidth | 0.64 deg |
| Gain | 48 dB |
| Receiver | |
| Noise Figure | 10 dB DSB |
| 1st IF | 1.4 GHz |
| 2nd IF | 160 MHz |
| Bandwidth | Variable, 5-40 MHz |
| Dynamic Range | 70 dB |

Range Capability

An approximate model for equivalent reflectivity factor, Z_e , of clouds and fog has been computed using Mie scattering theory and an appropriate particle distribution model. Z_e is used instead of Z to indicate non-Rayleigh scattering at this wavelength for particles larger than 100 micrometers. Using Deirmendjian's distributions several cloud types have been modeled at 1.4 mm (Ulaby, Moore and Fung, *Microwave Remote Sensing*, Vol. 1, pp. 290-309). Values for $\text{dB}(Z_e)$ are given in Table 2, along with extinction coefficient, K_e , for several cloud varieties and fog. Atmospheric loss due to water vapor, L_w , is typically 2 to 6 dB/km. Total atmospheric attenuation, L_a , can be approximated by $K_e + L_w$.

The radar range equation may be solved for Z_e using the parameters given and assuming 10 dB of integration gain (100 samples). Solving for the minimum Z_e as a function of range yields

$$\text{dB}(Z_e)_{\min} = 20 \log R + (L_a)R/500 - 80.2$$

where Z_e is the effective reflectivity factor in mm^6/m^3 , R is range in meters and L_a is atmospheric attenuation in dB/km. Figure 2 plots minimum $\text{dB}(Z_e)$ versus range for several atmospheric loss conditions. Also included is scattering cell volume versus range.

Figure 2 highlights the capability of this radar to detect very weak targets in comparatively small volumes. Applications of such an instrument include fine scale mapping of clear air - cloud boundaries and mapping of entire fair weather clouds or fog layers. Because of the extremely narrow beam (≈ 1 meter diameter at 100 meters range), this instrument is particularly attractive for fog measurements near ground level where reflections from terrain make such measurements difficult for centimeter wavelength radars. Periodic calibration targets along the radar path may also be used to estimate propagation losses in order to correct the measured reflectivity values.

Table 2. $\text{dB}(Z_e)$ and K_e for Various Cloud Varieties and Fog

| Cloud Type | $\text{dB}(Z_e)$ | K_e (dB/km) |
|-------------------|------------------|---------------|
| Stratocumulus | -8 | 2.8 |
| Fog Layer | -14 | 1.6 |
| Low-Lying Stratus | -16 | 2.6 |
| Cumulus Congestus | -14 | 22.5 |

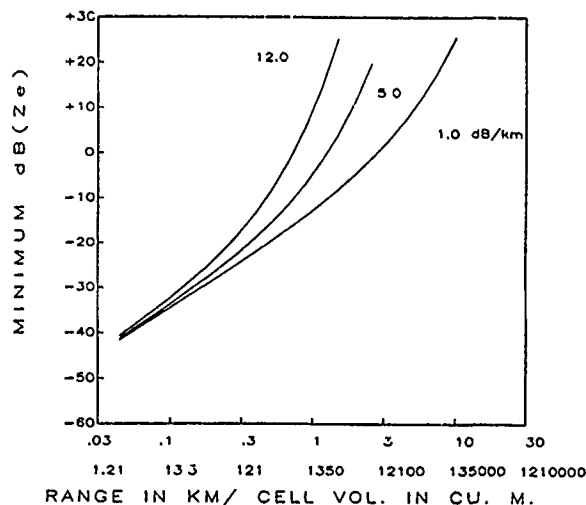


Figure 2. Minimum detectable reflectivity versus range for several atmospheric loss conditions. Range cell volume versus range is shown along abscissa.

Measurements

Measurements of fog and clouds have shown good agreement with predicted performance. Near horizontal measurements of fog with visibility varying between 300 and 800 meters were made over a period of one hour. A range vs. time diagram of fog reflectivity is shown in figure 3. Figure 4 is a map of stratocumulus reflectivity versus altitude over a period of 20 minutes. Significant backscatter was measured down to ground level, indicating the presence of small numbers of non-precipitating hydrometeors far below the apparent cloud base.

Conclusion

Both theory and preliminary measurements indicate the ability of this instrument to detect very weakly scattering clouds and fog over moderate ranges.

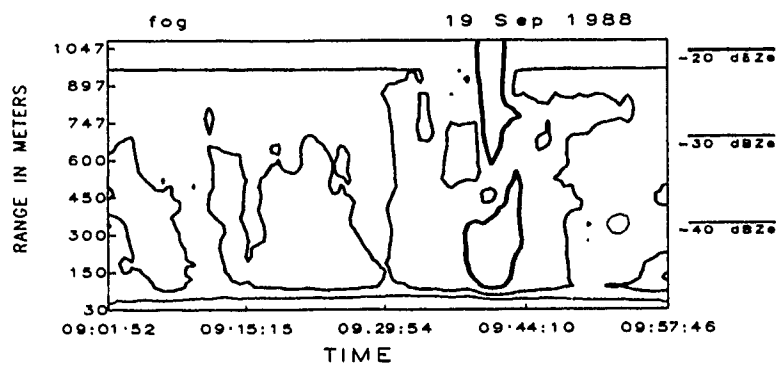


Figure 3. Range time diagram of fog reflectivity for near horizontal path.

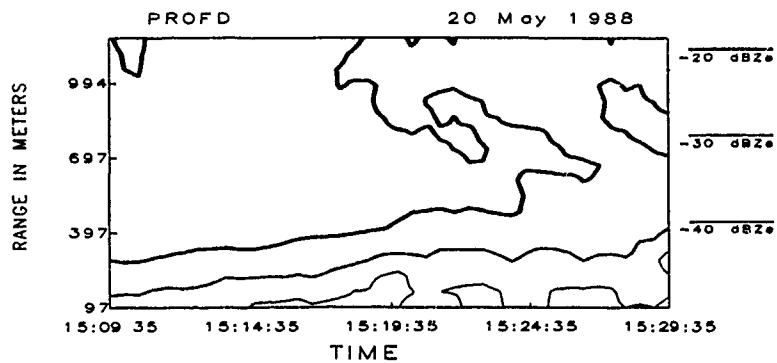


Figure 4. Range-time diagram of stratocumulus reflectivity for vertically pointed radar.

95 GHZ SHORT PULSE RADAR

E.A. ULIANA and B.S. YAPLEE
US Naval Research Laboratory
Washington D.C. 20375
(202)767-3443
N.C. CHU and M. NEWKIRK
Bendix Field Engineering
Oxon Hill, MD 20745

ABSTRACT

A two nanosecond, 95 GHz short pulse radar is being evaluated as an environmental remote sensor at the US Naval Research Laboratory. The initial tests of the radar utilized a wavetank where the water surface undulations can be controlled and monitored and the measured radar backscatter can be directly related to the surface manifestations. Measurements were made with wind velocities from no winds to 12 meter/sec, with look angles ranging from nadir to 50 degrees. Preliminary measurements indicate that for this short wavelength, the σ^0 curve does not appear to fall off as quickly. This paper will discuss the problems associated with using millimeter waves as an environmental remote sensor and some of the results obtained.

140 GHz SCATTEROMETER MEASUREMENTS

T. F. Haddock, F. T. Ulaby and J. East

Radiation Laboratory
University of Michigan
Ann Arbor, MI
USA

Abstract - The goal of the University of Michigan millimeter-wave radar program is to characterize terrain scattering at 35, 94 and 140 GHz. The 140 GHz channel of a truck-mounted scatterometer has recently been added to give the full desired operating capability. Two injection-locked 45.33-GHz Gunn oscillators use triplers to supply the up- and down-converters. Full polarization capability is obtained through the use of rotatable quarter-wave plates. Real-time signal processing and data reduction takes place in an HP 8510A automatic network analyzer on the truck-mounted platform. Sample measurements of the backscatter from vegetation and snow targets are given.

I. INTRODUCTION

The University of Michigan 140 GHz scatterometer system is the latest addition to the network-analyzer based millimeter-wave scatterometer system, a truck-mounted full-polarization scatterometer that has been developed in support of a program to characterize radar scattering from terrain at 35, 94, and 140 GHz. Operation of the scatterometer has been described elsewhere [1, 2]. Briefly, conversion from a swept 2 to 4 GHz IF to the millimeter-wave RF frequencies is made in the front-end, allowing flexible real-time signal processing by the remotely-located HP 8510A automatic network analyzer. An outline of the system is given in Fig. 1. The HP 8510 computer-control system allows vector error correction of system imperfections through its calibration algorithms. The system has previously operated in this mode at 35 and 94 GHz, and the 140 GHz channel is its latest extension in frequency capability.

II. 140-GHz SCATTEROMETER DESIGN

A block diagram of the 140-GHz front-end is shown in Fig. 2. The transmit portion across the top and the receive portion across the bottom are driven by a common local oscillator (LO) chain. The LO consists of a 45.33 GHz free-running Gunn oscillator, two circulator-coupled 45.33 GHz injection-locked Gunn oscillators acting as amplifiers and two times-3 frequency multipliers. This combination provides a nominal output power of 10 dBm from each multiplier. Other combinations of fundamental oscillators, amplifiers or frequency multipliers are possible. This particular combination provided the best combination of performance and cost. Polarization is set by a fixed quarter wave plate followed by a rotatable quarter-wave plate. A 90-degree rotation of the movable wave-plate moves the electric field vector through 90 degrees to give either vertical or horizontal polarization. The polarized signal is transmitted through a conical standard-gain horn with a half-power beamwidth of 11.8 degrees. The received RF signal passes through a 3.0 inch diameter lens-corrected horn antenna with a half-power beamwidth of 2.2 degrees. Receive

polarization is determined by movable and fixed quarter-wave plates, in the same manner as the transmit section. The RF signal is down-converted using a tripled 45.33 GHz LO. Since the conversion processes must be phase-coherent, the up- and down-converter LOs are each injection-locked from a central dual-ended Gunn oscillator running at 45.33 GHz. This arrangement gives a phase-coherent LO of sufficient power to supply both up- and down-converters.

III. CALIBRATION AND PERFORMANCE

For each data set, measurement of a sphere of known size and range is used to generate the 401 VV and HH calibration constants for each of the 401 frequencies in the 2 to 4 GHz IF band. VH and HV calibration constants are derived by measuring the response of a shorted delay-line on a standard-gain horn pointed at the radar. Its rectangular aperture when oriented at 45 degrees off horizontal gives equal V and H returns for a pure V or H incident signal. These measurements are used to establish the cross-polarization calibration constants. Sphere calibration is made on a daily basis, but the standard-gain horn cross-polarization calibration is more cumbersome and is made less frequently. Cross-polarization isolation of the system is typically about 15 dB, and this is checked at each system calibration by making cross-polarization measurements of the sphere. For most natural targets, the cross-polarized return at 140 GHz is only 3 to 6 dB below the like-polarized return. Hence the cross-polarization isolation of the system is quite adequate at 140 GHz. Noise performance of the system is checked after each calibration by making measurements of the sky at typical target ranges. Table I lists the measured system performance parameters.

TABLE I
Parameters of Truck-Mounted 140-GHz Scatterometer

| | |
|-------------------------------|---|
| RF Frequencies: | 138 to 140 GHz |
| Transmit power: | -4 dBm |
| RF Bandwidth: | 0 to 2.0 GHz |
| Sweep Rate: | 1 ms/freq., 51, 101, 201, 401 freq./sweep |
| Polarization: | VV, VH, HV, HH |
| Incidence Angles: | 0 to 70 degrees |
| Platform Height: | 2.7 meters minimum, to 18 meters maximum |
| Noise Equivalent σ^0 : | ~ -20 dB |

This work was supported by ARO contract DAAG 29-85-K-0220.

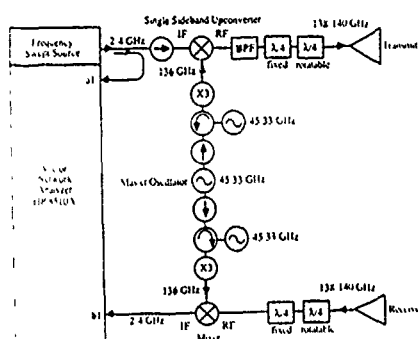


Fig. 2. Block diagram of 140-GHz scatterometer front-end.

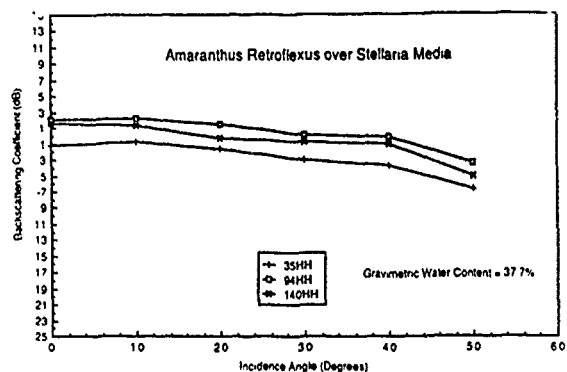


Fig. 5. Measurements of the same target as in Fig. 4 at all three millimeter-wave frequencies. Polarization is HH.

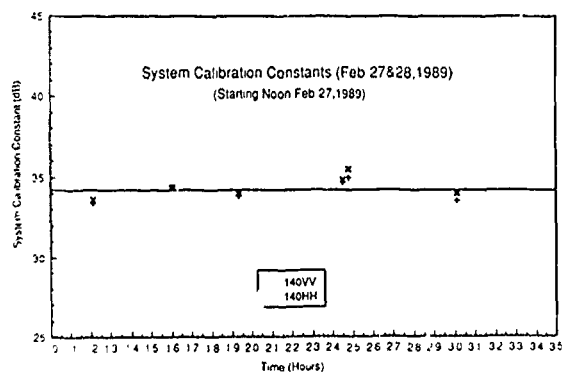


Fig. 3. Repeated calibrations of the system against a 15-inch diameter sphere over a diurnal cycle. The ordinate is time in hours, beginning with 10:00 AM, and the abscissa is the system calibration constant, averaged over 401 frequency points.

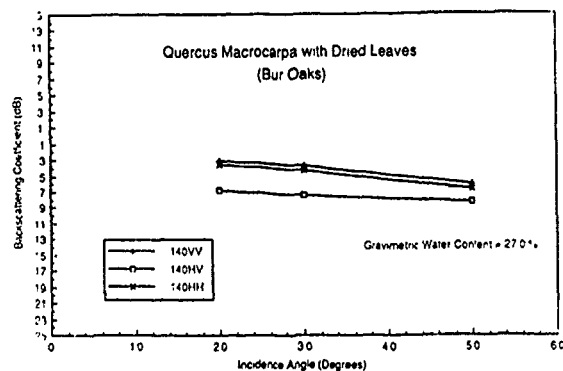


Fig. 6. Measurements of *Quercus Macrocarpa* trees (Bur Oaks) with dried leaves. Gravimetric water content 27.0%.

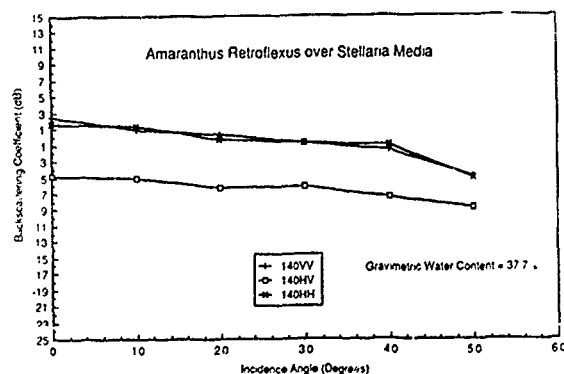


Fig. 4. Measurements of *Amaranthus Retroflexus* (Pigweed), about 50 centimeters high, over ground cover of *Stellaria Media* (Chick Weed). The gravimetric water content of the Pigweed was 37.7 %.

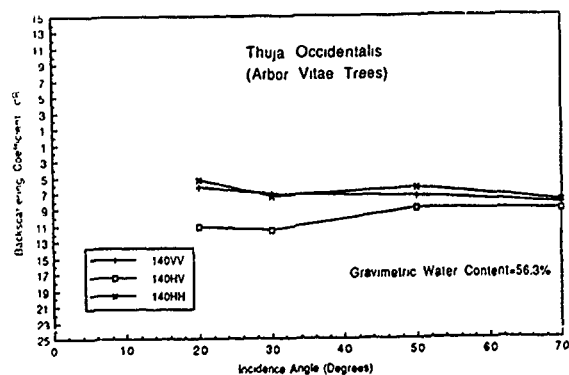


Fig. 7. Measurements of *Thuja Occidentalis* trees (Arbor Vitae) Gravimetric water content 56.3%.

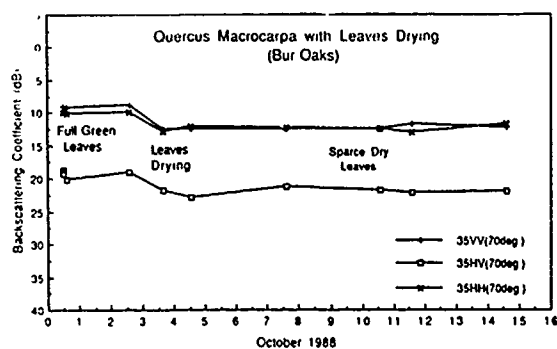


Fig. 8. 35 GHz measurements of the same target as in Fig. 6 as a function of time during the senescence of the leaves. Incidence angle is 70 degrees for all measurements.

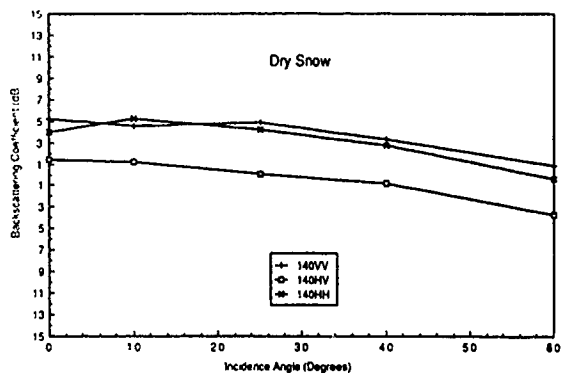


Fig. 9. Dry, metamorphosed snow with crystal size of approximately 2.2 millimeters.

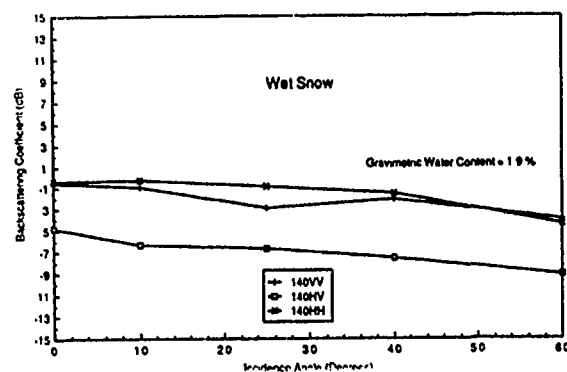


Fig. 10. Fresh, wet, unmetamorphosed snow, with crystal size of approximately 1.0 millimeter. Gravimetric water content 1.9 %

THE CRL MILLIMETER WAVE SCATTEROMETERS

Ken'ichi OKAMOTO, Toshio IHARA, Jun AWAKA, Hiromichi YAMASAKI, and Akira TAKAHASHI

Communications Research Laboratory, Ministry of Posts and Telecommunications,
2-1, Nukui-kitamachi 4-chome, Koganei-shi, Tokyo, 184, JAPAN

ABSTRACT

A groundbased millimeter wave scatterometer developed at CRL is described with preliminary experiments to measure backscatter of a soil called "Kanto" loam. The system is a continuous-wave type scatterometer operating at 51.35 GHz. The soil sample is spread uniformly, up to a depth of 30 cm, over an octagonal styrofoam box of about 2 meter diameter. This box rests on a turntable and, during the backscatter measurement, is rotated to get a large number of independent samples. As a preliminary experiment, the penetration depth of the soil was measured. We also have recently developed a 60 GHz polarimetric scatterometer which can measure not only the amplitude of scattering coefficient but also the phase difference between polarizations to construct the complete scattering matrix. Characteristics of the 60 GHz polarimeter are also given.

1. Introduction

A millimeter wave, short-distance radar can attain high spatial resolution using small size and light-weight instrumentation. Although millimeter waves are attenuated by rain, they can sense into fog, smog or dust where light waves cannot penetrate. Therefore, short-distance millimeter wave radars are expected to be used in various application fields such as collision avoidance sensing for automobiles, intrusion detection systems or level metering for a blast furnace and so on.

Data on millimeter wave scattering coefficients for various kinds of natural or artificial objects are, however, very sparse in comparison with those for microwave scattering. As a first step to make a data base of a various kinds of millimeter wave scattering coefficients, the Communications Research Laboratory (CRL) started to develop millimeter wave scatterometers and perform preliminary experiments to measure scattering coefficient of soil called "Kanto" loam, one of the typical soils in Japan. This paper briefly describes the CRL scatterometer and some results of preliminary experiments.

2. Preliminary Experiments by 50 GHz CW-scatterometer

2.1 50 GHz CW-scatterometer

The conceptual design of the total 50 GHz CW-scatterometer experiment system is shown in Fig. 1. It consists of a 50 GHz CW-scatterometer; a movable tower to carry the transmitter, receiver and antennas; an octagonal styrofoam sample box, a turntable to turn the sample box; control console for the movable tower and the turntable; data acquisition personal computer; and subsidiary instruments to measure the soil roughness, soil moisture content and the complex refractive index of soil. The characteristics of 50 GHz CW-scatterometer experimental system are given in Table 1. This is a bistatic radar system with pyramidal horn antennas. The distance between antennas and the targets is kept to more than 0.4 m to measure in the far field region of these antennas. The system's 51.35 GHz Gunn oscillator signal is radiated from the transmitter antenna to the target through the cross-coupling filter, and variable attenuator. The cross-coupling filter is introduced to change the transmitting polarization easily between the H and V polarizations. A cross coupling filter operates similarly in the receiver. The scattered wave is received and mixed with the local signal produced by doubling a Gunn oscillator signal of about 25 GHz to become 1F signal of 475 MHz. Finally, the 1F signal is then recorded as the received power through detector and log-amplifier. The maximum and minimum detectable powers are -25 dBm and -80 dBm, respectively and the dynamic range is about 55 dB.

Soil surface roughness and moisture content are two major parameters which affect backscatter measurements. We measure the backscatter while changing both soil moisture content and soil roughness. To maintain a quantitatively controllable target, backscatter measurements are performed indoors. The soil sample is spread uniformly, up to a depth of 30 cm, over the octagonal sample box of about 2 meter diameter. Samples are averaged to get scattering coefficients at fixed incidence angles. It is theoretically known that the number of independent samples obtained by one rotation of the turntable is approximately expressed by [1],[2]

$$N = 4\pi R \sin \theta / D \quad (1)$$

Here R is the range between the scatterometer and the target, θ is the incidence angle and D is the antenna aperture size. The received power data are recorded digitally on floppy disk after 12-bit A/D conversion with subsidiary data such as incidence angle, table rotation angle and so on. The statistical processing of data is performed off-line.

2.2 Calibration

To calibrate the instrument, a metal sphere of diameter 35 mm was placed in the beam at a distance of 1.11 m. Table 2 summarizes results from this experiment which agree with theory to better than 1 dB. During the experiment, the actual scattering angle was not 180° (complete back scattering) but 171° . Although uncertainty due to this effect has not yet been calculated, the error will be expected to be small.

2.3 Penetration depth of the soil

For the measurement of scattering coefficient of soil, an estimate of soil penetration depth is required. Therefore, an experiment to measure soil attenuation characteristics was performed. Figure 2 shows the arrangement of experiment. The scatterometer was set 1 m above the 1.1 m x 1.1 m flat aluminum plate placed horizontally in the sample box. The transmitter and receiver antenna beams were adjusted to satisfy a mirror reflection condition. Under this condition, the reflected power was measured and used as a reference for total reflection. A thin layer of "Kanto loam" soil was then spread over the aluminum plate. The reflected power was remeasured varying the thickness of the soil layer and referenced to the total reflection condition. The moisture content (weight) of soil was 46.5% and the density of soil was 0.71 g/cc in this experiment. The roughness of soil was estimated to be less than 0.5 mm, which is smaller than 1/8 of wave length of the 50 GHz scatterometer (0.58 mm), so the surface of the soil can be considered smooth from the Rayleigh roughness criterion.

In Fig. 3, measured values of reflectivity (expressed in dB) are shown by open circles, varying the thickness of the soil. The horizontal bars attached to the open circles indicate measurement errors in soil thickness. The reflectivity decreases as the thickness of soil increases when the layer is thin. However, the reflectivity value begins to oscillate when the thickness of soil becomes about 1 mm. This oscillating behavior can be interpreted as the reflected wave from the soil surface interfering with the reflected wave from the surface of aluminum plate. When the soil layer becomes thicker than about 10 mm, the reflectivity value seems to converge to be a constant value of -13 dB. This means the reflection from the aluminum plate has disappeared and reflection from soil surface has become dominant because of the large attenuation of millimeter waves

in soil.

The reflectivity under these conditions as shown in Fig. 3 can be approximated by an model, where the radio wave is vertically incident on an infinitely large metal plate with the layer of soil of constant thickness on it. The reflection coefficient can be approximated as [3]:

$$\gamma = \frac{\Gamma - \exp(-j2k_0nd)}{1 - \Gamma \exp(-j2k_0nd)} \quad (2)$$

Here, k_0 : wave number of radio wave in the air,

n : complex refractive index of soil ($n = n_1 - j \cdot n_2$),

Γ : reflection coefficient of soil,

d : thickness of soil layer.

The reflection coefficient of soil is related to the complex refractivity of soil by:

$$\left(n_1 - \frac{1 + |\Gamma|^2}{1 - |\Gamma|^2} \right)^2 + n_2^2 = \frac{4|\Gamma|^2}{(1 - |\Gamma|^2)^2} \quad (3)$$

The reflectivity $|\Gamma|^2$ of soil used in this experiment is about -13 dB as mentioned above. If we try to find value of n which satisfies equation (3) and which best fits in the measured reflectivity, n takes approximately the value

$$n = 1.52 - j \times 0.22 \quad (4)$$

Using this value of n , we calculate the attenuation coefficient A (dB/mm) as

$$A = 8.686 \times 2\pi/\lambda \times n_2 = 2.1 \text{ dB/mm} \quad (5)$$

The calculated reflectivity $|\gamma|^2$ and round trip attenuation $2 \cdot d \cdot A$ are shown in Fig. 3 as the dotted and solid line, respectively, using the estimated value of n . Although the calculated $|\gamma|^2$ and the measured reflectivity do not coincide in detail, we can say that the round trip attenuation value for soil with the thickness of 10 mm is approximately as large as 40 dB for the soil measured at the experiment. Therefore, a soil depth of 10 mm for the millimeter wave scattering is adequate unless the soil is very dry.

3. 60 GHz Polarimetric Scatterometer System

A 60 GHz polarimetric scatterometer has also been developed at CRL. This polarimetric scatterometer is based on the network analyzer and is similar to the millimeter wave polarimetric scatterometer developed by the University of Michigan [4]. The first measurements with this instrument will begin in the summer of 1989.

We show in Table 3 the characteristics of the 60 GHz polarimetric scatterometer.

This system can measure amplitudes and phase differences between HH-, HV-, VV- and VH-polarized scattering coefficients which compose all elements of scattering matrix. This is also a bistatic radar and can therefore measure not only radar backscatter coefficients but also scattering coefficients for the combination of any incidence and scattering angles. Scalar lens horn antennas with low sidelobe levels are used to reduce undesired signals from surrounding objects. A frequency interval of 1 GHz is swept stepwise between 57 GHz and 58 GHz and a range resolution of 15 cm can be attained.

4. Summary and Future Plans

The CRL millimeter wave 50 GHz CW and 60 GHz polarimetric scatterometer systems were introduced. The result of an end-to-end calibration experiment for the 50 GHz CW-scatterometer performed using a metal sphere of known radar cross section shows that the measured cross section coincided with the calculated value of the cross section within the error of 1dB. We show some results of a preliminary experiment to measure the attenuation coefficient of soil called "Kanto" loam. It was concluded that soils below a depth of 10 mm do not affect scattering measurements at 50 GHz or more for soils

with moderate to high moisture content.

We would like to continue in experiments to measure the scattering coefficient of "Kanto" loam as functions of incidence angle, surface roughness and soil moisture content using the CRL millimeter wave scatterometers. We also would like to measure the elements of scattering matrix for soil by the CRL polarimetric scatterometer. Finally we would like to add that we have a plan to develop millimeter wave polarimetric scatterometers at 80 GHz and 100 GHz.

References

- (1) Kobayashi, T. and H. Hisosawa, "Measurement of radar backscatter from rough soil surfaces using linear and circular polarizations", Int. J. Remote Sensing, 6, 2, pp.345-352 (1985)
- (2) M.I. Skolnik, "Radar Handbook", McGraw-Hill, Chap. 25, pp.11-16 (1970)
- (3) J.A. Stratton, "Electromagnetic Theory", McGraw-Hill, New York, p.512 (1941)
- (4) Ubiary F.T., T.F. Haddock, and M.E. Coluzzi, "Millimeter-wave bistatic radar measurements of sand and gravel", IGARSS'87, Univ. Michigan, Ann Arbor, pp.281-286 (1987)

Table 1 Characteristics of 50 GHz CW-scatterometer

| | |
|--------------------------|----------------------------------|
| <u>Scatterometer</u> | |
| Frequency | 51.35 GHz |
| Transmitter power | 12.2 dBm |
| Antennas | pyramidal horn antenna |
| Gain | 22.3 dB |
| Beamwidth (3dB) | 11.4° (E-plane), 13.2° (H-plane) |
| Polarization | HH, VV, HV, VH |
| Maximum detectable power | -25 dBm |
| Minimum detectable power | -80 dBm |
| <u>Tower</u> | |
| Incidence angle | 0 to 90° |
| Height | 0.8 to 4m |
| <u>Turntable</u> | |
| Rotation range | 0 to 360° (continuous) |
| Speed(variable) | 3 rpm(nominal) |

Table 2 Results of Calibration Experiment

| | |
|----------------------------------|------------------------|
| Frequency | 51.35 GHz |
| Transmitter power | 12.2 dBm |
| Transmitter antenna gain | 22.3 dBm |
| Receiver antenna gain | 22.3 dBm |
| Metal sphere radar cross section | -30.2 dBm ² |
| Distance | 1.11 m |
| Received power(calculated) | -52.8 dBm |
| Received power(measured) | -52 to -53 dBm |
| Background power(measured) | -75 dBm |

Table 3 Characteristics of 60 GHz Polarimeter

| | |
|-------------------|-----------------------------------|
| Frequency | 57-58 GHz |
| Transmitter power | 0 dBm |
| Antennas | scalar lens horn antenna |
| Gain | 35.5 dB |
| Beamwidth | 4.7° (E-plane), 4.6° (H-plane) |
| Polarization | HH, VV, HV, VH |
| Network analyzer | HP8753A |

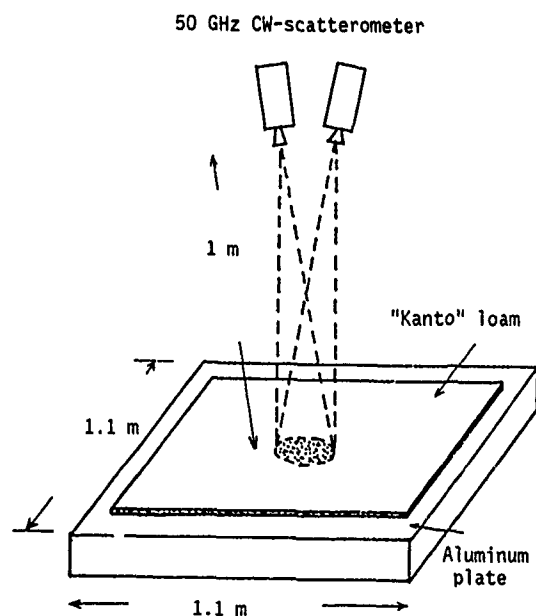


Fig. 2 Arrangement for Measurement of Soil Attenuation Characteristics

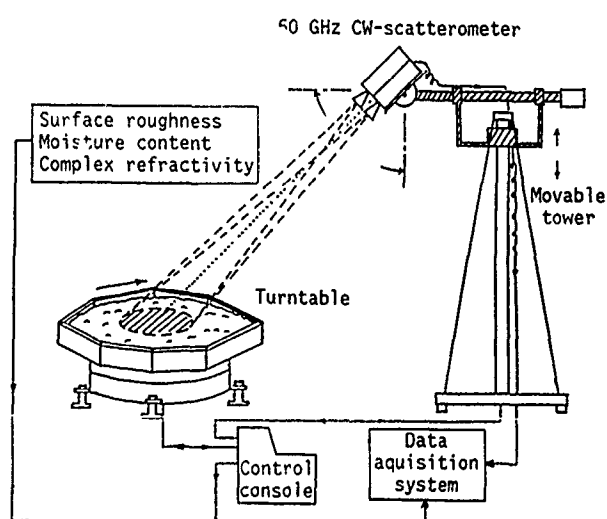


Fig. 1 Millimeter Wave Scattering System

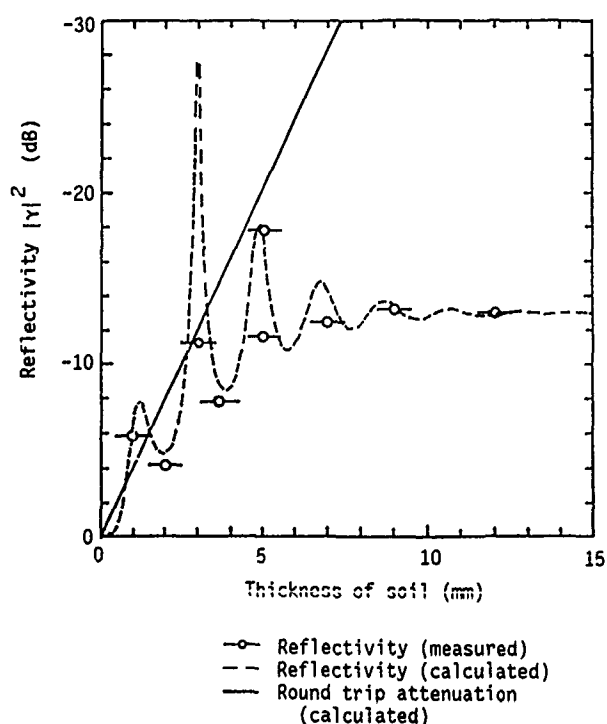


Fig. 3 Relation between the Thickness of Soil (mm) and the Reflectivity (dB)

MILLIMETER-WAVE PROPAGATION MEASUREMENTS
THROUGH CONIFER VEGETATION

D.L. Jones and R.H. Espeland
National Telecommunications and Information Administration
Institute for Telecommunication Sciences
325 Broadway
Boulder, Colorado 80303 USA
Telephone (303) 497-6295
FAX (303) 497-5993

Propagation measurements were made in the Olympic National Forest of Washington State during October, 1987 to examine millimeter-wave propagation through conifer vegetation. Linearly polarized CW signals at 9.6, 28.8, 57.6 and 96.1 GHz were used to evaluate attenuation, cross-polarization and back-scattering from conifer trees. Elevation and azimuth scans were conducted for various transmitter antenna heights and path lengths.

Very high attenuation rates (dB per meter of foliage) were measured on the shorter paths, while significantly lower rates were observed on paths extending deeper into the forest. Results from the measurements are presented and compared with data gathered from similar measurements taken through deciduous vegetation (F. K. Schwering, E. J. Violette and R. H. Espeland, "Millimeter-wave propagation in vegetation: experiments and theory", IEEE Trans. on Geoscience and Remote Sensing, vol. 26, no. 3, pp. 355-367, May 1988).

MILLIMETER WAVE SCATTERING MODEL FOR A LEAF

K. Sarabandi, F.T. Ulaby, and T.B.A. Senior

The Radiation Laboratory, EECS Department,
University of Michigan, Ann Arbor, MI 48109

Abstract

At millimeter wave frequencies a typical leaf is a significant fraction of a wavelength in thickness, and its nonuniform dielectric profile may affect its radar scattering. To provide a simple and efficient method for predicting the scattering, two types of physical-optics approximations are examined. The first approximates the volume polarization current by the current which would exist in an infinite dielectric slab with the same profile, while the second (and simpler) approach employs the surface current which, on the infinite slab, produces the known reflected field. It is shown that the first method is superior and predicts the scattered field to an accuracy which is adequate for most practical purposes.

1 Introduction

Leaves are a key feature of any vegetation canopy, and in order to model the scattering from vegetation-covered land, it is necessary to develop an efficient and effective technique for predicting the scattering from a single leaf. At microwave frequencies where a typical leaf is electrically thin with lateral dimensions at least comparable to the free space wavelength λ_0 , several methods have been proposed [e.g. Le Vine, et al. 1985, Willis, et al., 1988], all of which are based on the physical-optics approximation applied to a uniform dielectric slab. In particular, if the leaf thickness is no more than about $\lambda_0/50$, physical optics in conjunction with a resistive sheet model predicts the scattering at most angles of incidence [Senior, et al., 1987] and can also handle curved leaves [Sarabandi, et al., 1988].

On the other hand, at millimeter wavelengths the thickness can be a significant fraction of a wavelength, and it is also necessary to take into account the internal structure of a leaf. At least two different types of cell can be distinguished, and their differing water contents affect the dielectric constant, leading to a nonuniform dielectric profile. To compute the scattering at these higher frequencies, two different physical-optics approximations are examined. The first of these employs the polarization current which would exist in an infinite slab consisting of one, two or more layers simulating the dielectric profile of the leaf, and this is referred to as the volume integral physical optics (VIPO) approximation. The second (and simpler) approach postulates a surface current which, for an infinite slab, produces a plane wave identical to the reflected field, and this is the surface-current physical-optics (SCPO) approximation.

2 Structure of a Leaf

The structure of a typical vegetation leaf is shown in Fig.1. The type and number density of cells may vary as a function of depth into the leaf which, in turn, results in a nonuniform dielectric profile. The effect of this nonuniformity becomes observable at higher frequencies where the

thickness of the leaf is comparable to the wavelength.

Leaves contain two types of photosynthetic cells: *palisade parenchyma*, consisting of column-shaped cells in which most photosynthesis takes place, and *spongy parenchyma*, which consist of irregularly shaped cells with large spaces between them. Because a large part of the vegetation material is water, its dielectric constant is strongly influenced by the dielectric constant of water and the water content. For most leaves, the water content is higher in its upper layer (palisade region) than in the under surface (spongy region). The sensitivity of the dielectric constant of vegetation material to water content is much greater in the lower part of the millimeter-wave spectrum than in the upper part of the spectrum, but this is more than counterbalanced by the thickness to wavelength ratio. The net result is that the sensitivity to dielectric variations is greater at the higher frequencies.

To examine the effect of the nonuniform dielectric profile on the scattering properties of a leaf at millimeter wavelengths, we computed the normal incidence reflection coefficient Γ_0 of a two-layer dielectric slab and compared it with the reflection coefficient of a uniform dielectric slab whose dielectric constant is the average. The computation was performed for a leaf thickness of 0.5mm, and the water content ratio of the two layer was chosen to be 4 to 1, representing a marked variation between the upper and lower surfaces of the leaf. From the data in Table 1, it is seen that when the two-layer slab is approximated by a uniform slab the

| f (GHz) | ϵ_1 | ϵ_2 | $\epsilon_{(avg)} = \frac{\epsilon_1 + \epsilon_2}{2}$ | Γ_0 | $\Gamma_{0(avg)}$ |
|-----------|--------------|--------------|--|------------|-------------------|
| 35 | 20+i21 | 6+i3 | 13+i12 | 0.7426 | 0.782 - 0.16 |
| 94 | 6+i5 | 2+i1 | 4+i3 | 0.59212 | 0.4827 |
| 140 | 5+i4 | 2+i1 | 3.5+i2.5 | 0.50220 | 0.34226.1 |

Table 1: Voltage reflection coefficient for a two-layer and average dielectric slab

error in the reflection coefficient increases with increasing frequency, and is as large as 4 dB at 140 GHz.

3 Physical Optics Approximations

There are two types of physical optics approximation that can be employed. One is the surface current (SCPO) approach in which an infinite dielectric slab is replaced by an equivalent sheet current that produces a plane wave identical to the reflected wave of the slab. This current is then used as an approximation to the equivalent surface current over the upper surface of a finite dielectric plate. Alternatively, the induced (volume) polarization current in the plate can be approximated by the current in

the infinite dielectric slab, and we shall refer to this as the volume integral physical optics (VIPO) method. It is more accurate than the SCPO method, although the latter is more convenient to use for evaluating the scattered field.

To illustrate the two procedures, consider a dielectric plate consisting of a homogeneous dielectric of thickness d_1 and relative permittivity ϵ_1 atop a second material of thickness $d_2 - d_1$ and relative permittivity ϵ_2 . The plate occupies the region $-\frac{a}{2} \leq y \leq \frac{a}{2}$, $-\frac{b}{2} \leq z \leq \frac{b}{2}$, and $-d_2 \leq x \leq 0$ as shown in Fig. 2, and is illuminated by an E-polarized plane wave whose electric vector is

$$\mathbf{E}^i = \hat{y} e^{ik_0(x \sin \theta_0 - z \cos \theta_0)} \quad (1)$$

where k_0 is the propagation constant in the free space medium above and below the plate. When the plate is treated as an infinitely extended slab, the electric field can be written as

$$\begin{aligned} E_y &= (e^{-ik_0 z} + \Gamma e^{ik_0 z}) e^{ik_0 \sin \theta_0 x} & (0 \leq z) \\ E_y &= (B_1 e^{-ik_1 z} + A_1 e^{ik_1 z}) e^{ik_0 \sin \theta_0 x} & (-d_1 \leq z \leq 0) \\ E_y &= (B_2 e^{-ik_2 z} + A_2 e^{ik_2 z}) e^{ik_0 \sin \theta_0 x} & (-d_2 \leq z \leq -d_1) \\ E_y &= B_3 e^{-ik_0 z} e^{ik_0 \sin \theta_0 x} & (z \leq -d_2) \end{aligned} \quad (2)$$

where

$$k_{0x} = k_0 \cos \theta_0, \quad k_{jz} = k_0 \sqrt{\epsilon_j - \sin^2 \theta_0}$$

for $j = 1, 2$. If R_1 and R_2 are the reflection coefficients at the upper and lower surfaces where

$$R_1 = \frac{k_{0z} - k_{1z}}{k_{0z} + k_{1z}}, \quad R_2 = \frac{k_{0z} - k_{2z}}{k_{0z} + k_{2z}}$$

and

$$C_{\pm} = 1 \pm \frac{k_{2z}}{k_{1z}} \left\{ 1 + R_2 e^{2ik_{2z}(d_2 - d_1)} \right\} \left\{ 1 - R_2 e^{2ik_{2z}(d_2 - d_1)} \right\}^{-1},$$

application of the boundary condition at the three interfaces gives

$$B_1 = \frac{C_+(1+R_1)}{C_+ + C_- R_1 e^{2ik_{1z}d_1}}, \quad A_1 = \frac{C_-}{C_+} e^{2ik_{1z}d_1} B_1 \quad (3)$$

$$B_2 = \frac{e^{i(k_{1z} - k_{2z})d_1}}{1 - R_2 e^{2ik_{2z}(d_2 - d_1)}} \cdot \frac{2}{C_+} B_1, \quad A_2 = -R_2 e^{2ik_{2z}d_2} B_2 \quad (4)$$

$$B_3 = e^{i(k_{2z} - k_{0z})d_2} (1 - R_2) B_2$$

and

$$\Gamma = \frac{C_+ R_1 + C_- e^{2ik_{1z}d_1}}{C_+ + C_- R_1 e^{2ik_{1z}d_1}} \quad (5)$$

Given a volume distribution of electric current \mathbf{J} in free space, the corresponding Hertz vector in the far zone is

$$\Pi(\mathbf{r}) \approx \frac{e^{ik_0 r}}{k_0 r} \frac{iZ_0}{4\pi k_0} \int_V \mathbf{J}(\mathbf{r}') e^{-ik_0 \mathbf{r}' \cdot \mathbf{r}} dV, \quad (6)$$

and

$$\mathbf{E}(\mathbf{r}) \approx -k_0^2 \hat{\mathbf{r}} \times \hat{\mathbf{r}} \times \Pi(\mathbf{r}). \quad (7)$$

In the dielectric slab the volume current \mathbf{J} is the polarization current

$$\mathbf{J} = -ik_0 Y_0 (\epsilon_j - 1) E_y \hat{y}. \quad (8)$$

where E_y has the value appropriate to each layer ($j = 1, 2$), and when this is inserted into (6) and the integration carried out over the volume occupied by the plate, we obtain the VIPO approximation. For scattering in the direction θ_s indicated in Fig. 2 the expression for the Hertz vector is

$$\Pi^{VIPO} = \hat{y} \frac{e^{ik_0 r}}{k_0 r} \frac{k_0 ab \sin X}{4\pi X} F \quad (9)$$

where

$$\begin{aligned} F = & (\epsilon_1 - 1) \left\{ \frac{1 - e^{-i(k_{1z} - k_0 \cos \theta_s)d_1}}{i(k_{1z} - k_0 \cos \theta_s)} A_1 - \frac{1 - e^{-i(k_{1z} + k_0 \cos \theta_s)d_1}}{i(k_{1z} + k_0 \cos \theta_s)} B_1 \right\} \\ & + (\epsilon_2 - 1) \left\{ \frac{e^{-i(k_{2z} - k_0 \cos \theta_s)d_1} - e^{-i(k_{1z} - k_0 \cos \theta_s)d_2}}{i(k_{2z} - k_0 \cos \theta_s)} A_2 \right. \\ & \left. - \frac{e^{i(k_{2z} + k_0 \cos \theta_s)d_1} - e^{-i(k_{2z} + k_0 \cos \theta_s)d_2}}{i(k_{2z} + k_0 \cos \theta_s)} B_2 \right\} \end{aligned} \quad (10)$$

and

$$X = \frac{k_0 a}{2} (\sin \theta_s + \sin \theta_0). \quad (11)$$

The far zone scattered field can then be obtained from (7) and written as

$$\mathbf{E}^s = \frac{e^{ik_0 r}}{k_0 r} S_E(\theta_s, \theta_0) \quad (12)$$

where $S_E(\theta_s, \theta_0)$ is the far field amplitude, and for the VIPO approximation the result is

$$S_E^{VIPO}(\theta_s, \theta_0) = \hat{y} \frac{k_0^2 ab \sin X}{4\pi X} F. \quad (13)$$

In terms of the far field amplitude, the bistatic scattering cross section is

$$\sigma(\theta_s, \theta_0) = \frac{\lambda_0^2}{\pi} |S(\theta_s, \theta_0)|^2. \quad (14)$$

The SCPO approximation can be obtained by noting that the electric current sheet

$$\mathbf{J} = -2Y_0 \cos \theta_0 \Gamma e^{ik_0 \sin \theta_0 x} \delta(z) \hat{y} \quad (15)$$

produces a plane wave identical to the field reflected from the dielectric slab. When (15) is inserted into (6) we find

$$\Pi^{SCPO} \approx \hat{y} \frac{e^{ik_0 r}}{k_0 r} \cdot \frac{-i}{2\pi} \cos \theta_0 \Gamma ab \frac{\sin X}{X}, \quad (16)$$

and the far field amplitude is then

$$S_E^{SCPO}(\theta_s, \theta_0) = \hat{y} \frac{-ik_0^2}{2\pi} \cos \theta_0 \Gamma ab \frac{\sin X}{X}. \quad (17)$$

In the specular ($\theta_s = -\theta_0$) and backscattering ($\theta_s = \theta_0$) directions it can be verified that (13) and (17) are identical, but in the other directions the two approximations differ.

In the case of H polarization for which

$$\mathbf{H}^i = \hat{y} e^{ik_0(x \sin \theta_0 - z \cos \theta_0)} \quad (18)$$

the analysis is similar. With H_y represented as shown in (2), the various coefficients (now indicated by primes) differ from those for E polarization in having k_{1z} replaced by k_{1z}/ϵ_1 and k_{2z} replaced by k_{2z}/ϵ_2 everywhere except in the exponents. The induced polarization current then has two components and is given by

$$\mathbf{J} = -ik_0 Z_0 (\epsilon_j - 1) (E_x \hat{x} + E_z \hat{z}). \quad (19)$$

where $E_x = (ik_0 \epsilon_j)^{-1} Z_0 \partial H_y / \partial z$ and $E_z = -(ik_0 \epsilon_j)^{-1} Z_0 \partial H_y / \partial x$ have the values appropriate to each layer ($j = 1, 2$). The Hertz vector can be computed using (6), and for the scattered field \mathbf{H}^s , the far field amplitude is found to be

$$S_H^{VIPO}(\theta_s, \theta_0) = \hat{y} \frac{k_0^2 ab \sin X}{4\pi X} (\cos \theta_s F'_1 - \sin \theta_s F'_2), \quad (20)$$

where

$$\begin{aligned} F'_1 = & \left\{ \frac{k_{1z}(\epsilon_1 - 1)}{k_{1z}} \left\{ \frac{1 - e^{-i(k_{1z} - k_0 \cos \theta_s)d_1}}{i(k_{1z} - k_0 \cos \theta_s)} A'_1 + \frac{1 - e^{-i(k_{1z} + k_0 \cos \theta_s)d_1}}{i(k_{1z} + k_0 \cos \theta_s)} B'_1 \right\} \right. \\ & + \frac{k_{2z}(\epsilon_2 - 1)}{k_0 \epsilon_2} \left\{ \frac{e^{-i(k_{2z} - k_0 \cos \theta_s)d_1} - e^{-i(k_{1z} - k_0 \cos \theta_s)d_2}}{i(k_{2z} - k_0 \cos \theta_s)} A'_2 \right. \\ & \left. \left. + \frac{e^{i(k_{2z} + k_0 \cos \theta_s)d_1} - e^{-i(k_{2z} + k_0 \cos \theta_s)d_2}}{i(k_{2z} + k_0 \cos \theta_s)} B'_2 \right\} \right\}, \end{aligned} \quad (21)$$

and

$$F_2' = \sin \theta_0 \left\{ \frac{\epsilon_1 - 1}{\epsilon_1} \left[\frac{1 - e^{-i(k_{1z} - k_0 \cos \theta_0) d_1}}{i(k_{1z} - k_0 \cos \theta_0)} A_1' - \frac{1 - e^{i(k_{1z} + k_0 \cos \theta_0) d_1}}{i(k_{1z} + k_0 \cos \theta_0)} B_1' \right] \right. \\ \left. + \frac{\epsilon_2 - 1}{\epsilon_2} \left[\frac{e^{-i(k_{2z} - k_0 \cos \theta_0) d_1} - e^{-i(k_{1z} - k_0 \cos \theta_0) d_2}}{i(k_{2z} - k_0 \cos \theta_0)} A_2' \right. \right. \\ \left. \left. - \frac{e^{i(k_{2z} + k_0 \cos \theta_0) d_1} - e^{i(k_{1z} + k_0 \cos \theta_0) d_2}}{i(k_{2z} + k_0 \cos \theta_0)} B_2' \right] \right\}. \quad (22)$$

The SCPO approximation can also be obtained by noting that a magnetic current sheet of the form

$$J^* = -2Z_0 \cos \theta_0 \Gamma' e^{ik_0 \sin \theta_0 x} \delta(z) \hat{y} \quad (23)$$

generates a plane wave identical to the reflected wave. Using this as the equivalent surface current on the dielectric plate, the magnetic far field amplitude becomes

$$S_H^{SCPO}(\theta_s, \theta_0) = \hat{y} \frac{-ik_0^2}{2\pi} \cos \theta_0 \Gamma' ab \frac{\sin X}{X}. \quad (24)$$

As in the case of E polarization, the two approximations are identical in the specular direction, but (20) and (24) differ in all other directions including backscattering ($\theta_s = \theta_0$) unless $\theta_0 = 0$.

4 Numerical Results

To illustrate the difference between the VIPO and SCPO approximations and to test their accuracy the two approximations have been compared with the results of a moment method solution of the volume integral equation. For a $2\lambda_0$ square plate of thickness $d_2 = \lambda_0/4$ with $\epsilon_2 = \epsilon_1 = 3 + i0.1$ and illuminated by an E-polarized plane wave at normal incidence, the two approximations are compared with the moment method solution in Fig. 3, and the superiority of VIPO is clear.

In the case of a thin plate the two approximations are indistinguishable. This is illustrated in Fig. 4 showing the VIPO expression (13) and the moment method solution for a $2\lambda_0$ square plate of thickness $d_2 = \lambda_0/50$ for E polarization. The plate is a homogeneous one having $\epsilon = 13 + i12$ corresponding to the average permittivity at 35 GHz in Table 1. The SCPO expression (18) yields the same results, as does a two-layer model having the permittivities listed in Table 1. Over a wide range of scattering angles, the approximate and moment method solutions are in excellent agreement.

As the frequency and, hence, the electrical thickness of the plate increase, the superiority of the VIPO approximation becomes apparent and, in addition, it becomes necessary to take the layering of the plate into account. In Figs. 5 and 6 the simulated frequency is 140 GHz, but to keep the moment method calculations tractable, the plate has been reduced in size to $1.4\lambda_0$ by $2\lambda_0$. The curves shown are for a two-layer plate having $d_2 = 2d_1 = 0.5mm$ with $\epsilon_1 = 5 + i4$ and $\epsilon_2 = 2 + i1$, and for a single layer having the average permittivity $\epsilon_{avg} = 3.5 + i2.5$ (see Table 1). Since the accuracy of the physical optics approximation increases with the plate size, the agreement between the two-layer VIPO approximation and the moment method solution is remarkably good, and significantly better than if a single layer had been used.

5 Conclusions

A typical leaf has at least two dielectric layers whose cells have differing water content, and this produces a nonuniform dielectric profile which can now affect the scattering. At microwave frequencies where the leaf is no more than (about) $\lambda_0/50$ in thickness, the nonuniformity is not important, and as shown by Senior, et al. [1987] the leaf can be modeled as a resistive sheet using an average value for the permittivity. At higher frequencies, however, the thickness and structure of a leaf are more significant. At 100 GHz and above a leaf is a considerable fraction of a wavelength in thickness, and in spite of the reduced sensitivity to water content, the nonuniformity affects the scattering.

For a two-layer model of a leaf, the SCPO approximation has been compared with the volume integral (VIPO) approximation. When the leaf is thin the two approximations are identical and in good agreement with data obtained from a moment method solution of the integral equation.

but as the electrical thickness increases, the two approximations diverge in all directions except the specular and (for E polarization) backscattering directions. Although the VIPO approximation is more complicated, its accuracy is greater, and the agreement with the moment method data is better using a two-layer model than when a single layer of average permittivity is employed.

For most practical purposes it would appear that VIPO in conjunction with an accurate dielectric profile of a leaf provides an adequate approximation to the scattering at millimeter wavelengths. As our knowledge of the profile increases, it may be desirable to use a multi-layer model which could even simulate a continuous, nonuniform profile, and a convenient way of doing this is described by Sarabandi, et al. [1989]. We also note that at frequencies for which the leaf thickness is comparable to $\lambda_m/2$ where λ_m is the (average) wavelength in the leaf, the scattering is greatly reduced at some angle of incidence, and because the permittivity is complex, there is actually a range of angles for which this is true. Since the reduction is accompanied by an increase in the field transmitted through the leaf, this could provide a means for penetration through a vegetation canopy.

Acknowledgement

This work was supported by the U.S. Army Research Office under contract DAAG 29-85-k-0220.

References

- [1] Le Vine, D.M., A. Snyder, R.H. Lang, and H.G. Garter, Scattering from thin dielectric disks, *IEEE Trans. Antennas Propag.*, 33, 1410-1413, 1985.
- [2] Sarabandi, K., T.B.A. Senior, and F.T. Ulaby, Effect of curvature on the backscattering from a leaf, *J. Electromag. Waves and Applics.*, 2, 653-670, 1988.
- [3] Sarabandi, K., F.T. Ulaby, and T.B.A. Senior, Millimeter wave scattering model for a leaf, submitted to *Radio Sci.* for publication.
- [4] Senior, T.B.A., K. Sarabandi, and F.T. Ulaby, Measuring and modeling the backscattering cross section of a leaf, *Radio Sci.*, 22, 1109-1116, 1987.
- [5] Senior, T.B.A., and J.L. Volakis, Sheet simulation of a thin dielectric layer, *Radio Sci.*, 22, 1261-1272, 1987.
- [6] Willis, T.M., H. Weil, and D.M. Le Vine, Applicability of physical optics thin plate scattering formulas for remote sensing, *IEEE Trans. Geosci. Remote Sensing*, 26, 153-160, 1988.

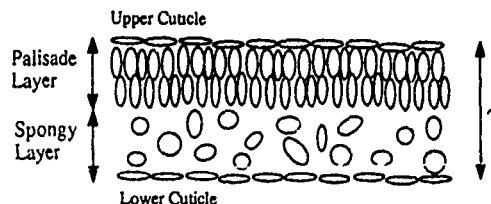


Figure 1: The structure of a typical vegetation leaf.

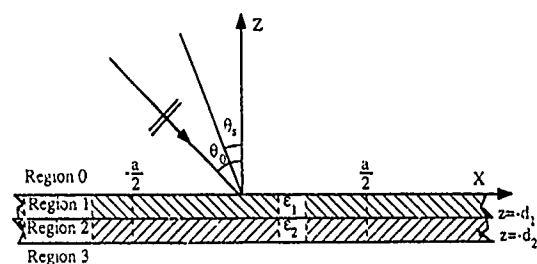


Figure 2: The geometry of the scattering of a plane wave from a two-layer dielectric slab.

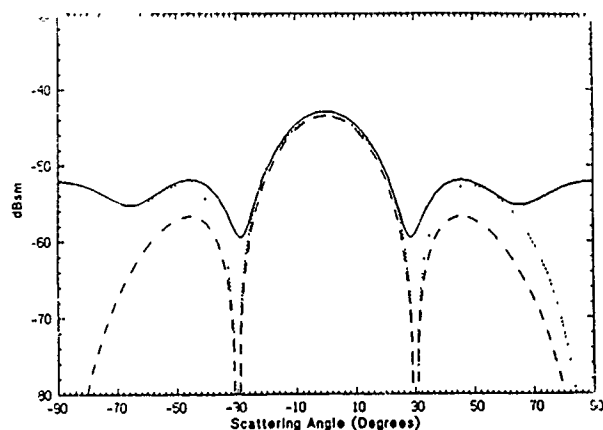


Figure 3: The bistatic cross section of a $2\lambda_0 \times 2\lambda_0$ plate for E polarization with $d_2 = \lambda_0/4$ and $\epsilon_1 = \epsilon_2 = 3 + i0.1$ at normal incidence: (—) moment method solution, (---) VIPO, (- -) SCPO.

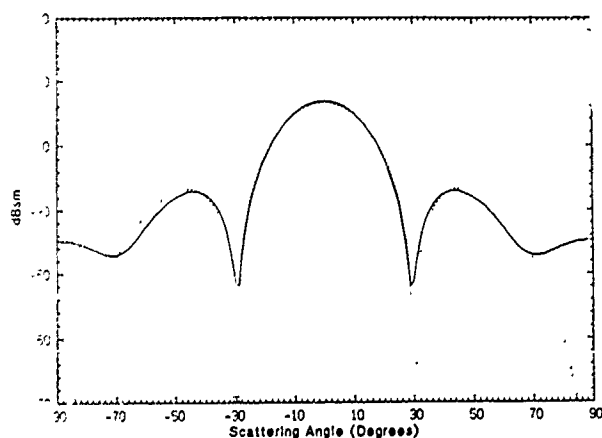


Figure 4: The bistatic cross section area of a $2\lambda_0 \times 2\lambda_0$ plate for E polarization with $d_2 = \lambda_0/50$ and $\epsilon_{avg} = 13 + i12$ at normal incidence: (—) moment method solution, (---) VIPO or SCPO.

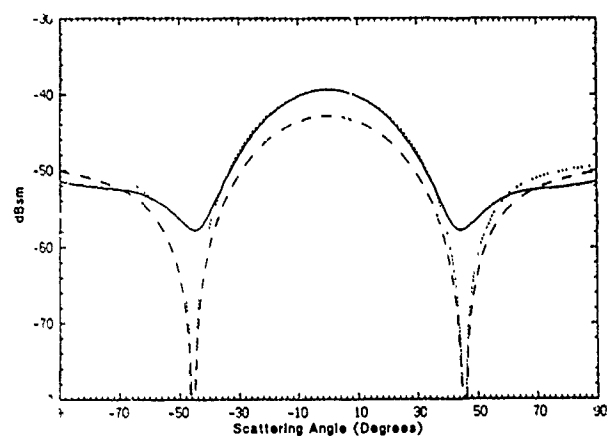


Figure 5: The bistatic cross section of a $1.4\lambda_0 \times 2\lambda_0$ plate for E polarization with $d_2 = 2d_1 = 0.5mm$ and $f = 140$ GHz at normal incidence: (—) moment method solution with $\epsilon_1 = 5 + i4$, $\epsilon_2 = 2 + i1$, (---) VIPO with $\epsilon_1 = 5 + i4$, $\epsilon_2 = 2 + i1$, (- -) VIPO with $\epsilon_2 = \epsilon_1 = 3.5 + i2.5$.

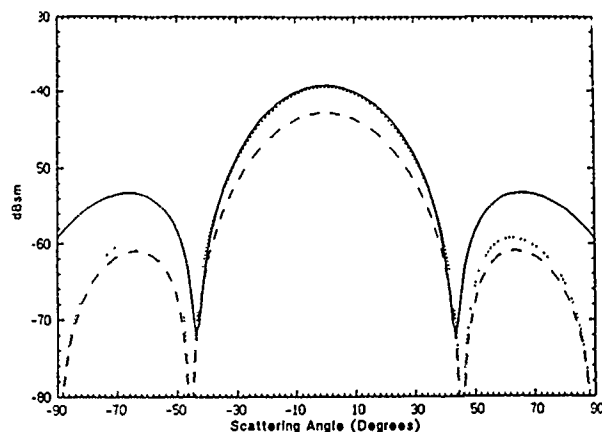


Figure 6: The bistatic cross section area of a $1.4\lambda_0 \times 2\lambda_0$ plate for H polarization with $d_2 = 2d_1 = 0.5mm$ and $f = 140$ GHz at normal incidence: (—) moment method solution with $\epsilon_1 = 5 + i4$, $\epsilon_2 = 2 + i1$, (---) VIPO with $\epsilon_1 = 5 + i4$, $\epsilon_2 = 2 + i1$, (- -) VIPO with $\epsilon_2 = \epsilon_1 = 3.5 + i2.5$.

MEASUREMENTS AND MODELING OF
MILLIMETER-WAVE SCATTERING FROM TREE CANOPIES

F. W. Ulaby, T. F. Haddock, and Y. Kuga
The Radiation Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109-2122
Telephone: (313) 764-0500
FAX: 313/936-3492
TELEX: 432-0815 UOFM-UI

ABSTRACT

Using a truck-mounted scatterometer system, measurements were conducted to study the backscatter response of tree canopies at 35, 94, and 140 GHz. The measurements were made as a function of incidence angle at HH, VV, and HV polarizations for several types of trees and tree conditions. A second-order radiative transfer model was developed using an azimuthically symmetric phase function characterized by a narrow-lobed forward scattering pattern superimposed on an isotropic background pattern. This approach yields a computationally simple model that provides good agreement with the experimented data for both like- and cross-polarization.

SPACEBORNE TECHNOLOGY CONTRIBUTIONS TO HYDROLOGICAL STUDIES IN THE CONTEXT OF GLOBAL CHANGE

Vincent V. Salomonson

NASA/Goddard Space Flight Center
Space and Earth Sciences Directorate
Greenbelt, MD 20771
U.S.A.

ABSTRACT

Spaceborne observations have become increasingly accepted for studies of the hydrological cycle and for information that can be used in better management or assessment of water resources over large areas. As data records are extended over long periods of time using sensors that improve in their reliability and quality including radiometric calibration, the utility of spaceborne observations will become increasingly evident. The Earth Observing System (Eos) concept being studied by NASA offers some real potential for providing improved means and procedures for hydrological studies, particularly on a global basis, employing spaceborne sensors and a comprehensive information system. Continuing emphasis needs to be placed on research exploring the interpretation of remotely sensed observations through the assimilation and use of these data in models of hydrological processes.

Key Words: hydrology, sensors, Earth Observing System, water resources management, data sets.

1. INTRODUCTION

One of the central problems in the earth sciences is the question of climate change. The question not only revolves around the mechanisms of climate change and when and at what magnitude it is occurring or will occur, but also the effects that anthropogenic activities may have in causing climate change. In the latter instance, it is becoming increasingly clear that man is causing large changes to occur on the surface of the Earth and in its atmosphere. It remains to be seen what these changes mean in terms of providing an external forcing that will change the climate and cause a subsequent redistribution of resources upon which so much life and the industrial, agricultural, and recreational activities of man depend. The abundance of water is an aspect which distinguishes the Earth from other planets. However, our knowledge of the spatial and temporal distribution of the water on the planet and the mechanisms which cause the differences in this distribution is significantly lacking. It is imperative that the hydrological cycle and its relationship to climate change be understood in order to appreciate the impact on the water resources of the Earth.

The use of remote sensing from aircraft and spacecraft has progressed sufficiently so that it is generally acknowledged that such capabilities can and should be a component of any attempts to study global hydrological processes. This paper will focus on the use of spacecraft observations for such purposes including not only the advances that have been made up to the present, but what may be realistically expected in the future, perhaps by the end of this century. In the latter instance, the focus will be on NASA plans for satellite remote sensing as it relates to the monitoring of the parameters comprising the hydrological cycle and the water resources of the Earth.

2. A BRIEF SUMMARY OF KEY ACCOMPLISHMENTS TO DATE

Spaceborne sensors have already made key contributions that establish the maturing of such observations for hydrological studies. In the most elemental case, spaceborne observations can be used to observe the extent of hydrological features such as snow and ice cover, vegetation cover, etc., and the change in the extent of these features over time. As time has gone by, there has been an increasing capability to extract estimates of volume, rate or condition of hydrologically related features from remotely sensed measurements. Examples of such measurements include surface temperature, snow depth, or monthly rainfall estimates.

Observations such as those just suggested have come, most notably, from sensors on the meteorological satellite series operated by the National Oceanic and Atmospheric Administration (NOAA), the Landsat satellite series operated by NASA starting in 1972 and later by NOAA and the EOSAT Corporation, and the Nimbus satellite series operated by NASA. In the last instance, the most prominent example comes from the Nimbus-7 satellite which has operated for approximately one decade starting in 1979. Other satellites and their attendant sensors are starting to complement the aforementioned systems. Examples are the SPOT satellite series operated by France and the Meteosat series operated by the European Space Agency (ESA).

As these sensors compile consistent data sets over a number of years, the attendant observations become increasingly valuable with time. For example, observations of snowcover have been obtained in nearly continuous fashion by optical television or radiometric sensors on the NOAA satellites since 1966 and, as limited only by the evolving capability of the sensors and the techniques used to analyze and archive the data, serve as a data set to examine to see if climate change is reflected [1].

At present the key NOAA sensors are the Advanced Very High Resolution Radiometer (AVHRR) and the Tiros Operational Vertical Sounder (TOVS) system consisting of the High Resolution Infrared Sounder (HIRS) and the Microwave Sounding Unit. These systems have gathered observations of useful, radiometric quality since 1979 or approximately one decade. Besides snow and ice cover, the extent of land cover types such as various kinds of vegetation cover [2] and cloudiness (which is related to precipitation) can be observed from the AVHRR. The TOVS observations, again extending from 1979, depict changes in land surface and ocean surface temperature, cloudiness and radiation balance components at the top of the atmosphere that are related to the dynamics of the hydrological cycle [3]. Selected high spatial resolution, regional observations from the Landsat sensors over climate sensitive regions or features such as glaciers extend back to 1972. The Nimbus-7 satellite and the observations extending over nearly a decade from the Scanning Multichannel Microwave Radiometer (SMMR) appear to be

providing very useful observations of snow extent, snow depth and wetness variability [4] and land cover or soil wetness changes that complement the optical sensor observations such as those from the AVHRR [5].

In essence, it can be concluded that observations from satellites can offer assured, dependable observations of the extent of land surface cover types such as bare soil, general vegetation cover types such as forest, grassland, and cropped areas plus the snow and ice cover and regions covered by water. Satellite sensors have shown considerable potential for observing hydrologically related conditions or states such as surface temperature, components of the radiation balance at the surface or at the top of the atmosphere, and cloudiness as related to precipitation, but much remains to be done to show how these observations can be routinely used in models of the general circulation of the atmosphere, including the hydrologic cycle, and in process or watershed models. Reviews of the progress and state of advancement in these areas are available [6,7,8]. As the length of record increases, the opportunity to develop insights as to the geographical extent and duration of climatological conditions related to hydrological conditions increases. An example of such a situation is nicely described in conjunction with tropical cloudiness variations and subsequent effects on wet and dry conditions over large areas due to teleconnections in the atmosphere [9,10]. This simply reinforces the point that spaceborne remote sensing observations increase in their utility and value as they are consistently and carefully acquired and archived over increasingly long periods of time.

3. THE EARTH OBSERVING SYSTEM (Eos)

The Eos [11] is a system employing many instruments to obtain global observations that will be of use in extending observations such as those noted in section 2. The Eos systems are being designed to operate for 10-15 years and thus provide long-term data sets that can be used for scientific and resource management studies. Key among the sensors being planned for Eos are the so-called facility instruments. The facility instruments that easily relate to hydrological studies are those listed below:

- Moderate Resolution Imaging Spectrometer (MODIS)
- Advanced Microwave Scanning Radiometer (AMSR)
- Atmospheric Infrared Sounder (AIRS)
- Advanced Microwave Sounding Unit (AMSU)
- High-Resolution Imaging Spectrometer (HIRIS)
- Synthetic Aperture Radar (SAR).

The instruments noted above will extend data bases referred to in section 2 and enhance the possibility for better observing hydrologically relevant parameters. The MODIS will extend the data record of the present AVHRR and be very useful for global snow and ice mapping, vegetation cover mapping, etc. The AMSR will extend and enhance the capabilities presently represented by the Nimbus SMMR and the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I). The AIRS and the AMSU represent a much improved capability over the present TOVS system. Finally, the HIRIS and the SAR, in essence, extend greatly the capabilities for high-resolution, regional surveys of surface cover conditions as now provided over extended periods by the Landsat Thematic Mapper and the Multispectral Scanner (MSS) by providing spectral coverage spanning the electromagnetic spectrum from the visible to the microwave region. Other facility instruments such as the Geoscience Laser Ranging System (GLRS) and the radar altimeter (ALT) may also serve hydrological studies through providing more accurate and widely available topographic information for hillslope hydrology studies, as an example [12]. Some other principal investigator instruments under consideration for flight and operation on the Eos may also provide improved, important information about components of the radiation balance that drive hydrometeorological processes such as evapotranspiration.

4. BEYOND AND IN ADDITION TO Eos

There is much that needs to be accomplished in addition to Eos in order to fully understand the hydrological cycle of the Earth and to make greater use of the advantages of remotely sensed observations from spaceborne platforms. An example of a space mission that is needed to complement the Eos is the proposed Tropical Rainfall Measuring Mission (TRMM) [13]. This mission would use a non-sun synchronous orbit and radar and passive microwave sensors to make a better sampling of the diurnal variability of rainfall and subsequently provide a better understanding of the role of latent heat release in the tropical regions. Focusing on this subject in the tropical latitudes will address some of the primary processes and dynamics that drive the dynamics of the hydrological cycle.

In a similar vein, in studies being led and managed at the Marshall Space Flight Center of NASA, advancements in geosynchronous satellite systems are being considered. These advancements would be put into place on geosynchronous platforms presently being scheduled to operate at the very end of this century or very soon after the year 2000. Key among the advancements being considered is the inclusion of passive microwave observations. This would undoubtedly provide some new observations that would offer new insight using the high temporal sampling capability offered by a geosynchronous orbit. Key among the hydrological parameters that would be observed is rainfall.

It is imperative that well-planned and focused programs be organized to study hydrological processes and better understand and utilize spaceborne remotely sensed observations. An exemplary program being planned is the Global Energy and Water Experiment (GEWEX) [14] presently scheduled for the late 1990's, approximately concurrent with the TRMM and Eos missions. This experiment focuses, as can be overtly deduced from the title of the experiment, on hydrological processes and strongly coupled energy balance parameters. This, additionally, is quite an appropriate subject to address in conjunction with satellite observations in that it focuses on the strength of remotely sensed observations in terms of their fundamentally measuring radiance and fluxes emerging from the top of the earth-atmosphere system. Another focus and noteworthy aspect of this effort is the focus on models of hydrological processes and the importance of studying how remotely sensed observations can be best assimilated, tested and studied, and, subsequently, better utilized or exploited for global meteorological and hydrological studies.

5. SUMMARY AND CONCLUSIONS

Many advances have been made in the use of satellite observations for the study of the hydrological cycle. The major point made in this paper is that a key advancement is the accumulation of global satellite observations over several years. These observations offer improved insight as to the monthly, seasonal and interannual variability of hydrological phenomena along with the nearly synoptic, large area to global coverage that is advantageously offered by spaceborne sensors. Examples are the global extent of snow and ice, cloudiness as it relates to precipitation, and vegetation dynamics that relate to the presence of surface moisture and evapotranspiration.

The Eos being planned for the late 1990's should very appropriately extend and improve the data bases being accumulated by existing satellite systems. The Eos Data and Information System (EosDIS) will be given the appropriate attention so that the Eos sensor data can be effectively utilized. Furthermore, complementary technology efforts such as the TRMM mission concept and modeling studies having the objective of improving models so as to take advantage of remotely sensed observations need to be pursued and encouraged. In the latter instance, focused study efforts like the GEWEX are important in order to more completely understand the global hydrological cycle and gain insights about the availability and distribution of the water resources of the world.

6. ACKNOWLEDGMENTS

The author wishes to thank especially Kelly Pecnick for expertly formatting and preparing this paper for publication. He also wishes to thank his colleagues at the Goddard Space Flight Center for the help they have provided through discussions and the provision of references and other insights that have contributed to this paper.

7. REFERENCES

1. Wiesnet, D. R., C. F. Ropelewski, G. J. Kukla, and D. A. Robinson, "A discussion of the accuracy of NOAA satellite-derived global seasonal snow cover measurements," *Proceedings of the Symposium on the Large Scale Effects of Snowcover* (Edited by B. E. Goodison, R. G. Barry, and J. Dozier), IAHS Publication No. 166, 291-304, 1987.
2. Justice, C. O. (Editor), "Monitoring the grasslands of semi-arid Africa using NOAA-AVHRR data," *International Journal of Remote Sensing*, Vol. 7, pp. 1383-1622 (special issue), 1986.
3. Susskind, J., J. Rosenfield, D. Reuter, and M. T. Chahine, "Remote sensing of weather and climate parameters from HIRS-MSU on TIROS-N," *Journal of Geophysical Research*, Vol. 89, pp. 4677-4697, 1984.
4. Foster, J. L., D. K. Hall, and A.T.C. Chang, "Remote sensing of Snow," *EOS, Transactions of the American Geophysical Union*, Vol. 68, pp. 681-684, 1987.
5. Choudhury, B. J., "A new long-term data set for biospheric studies," *International Journal of Remote Sensing*, Vol. 9, No. 2, pp. 185-186, 1988.
6. Salomonson, V. V., et al., "Water Resources Assessment," *Manual of Remote Sensing*, Volume II (Editors: J. E. Estes and G. A. Thorley), Chapter 29, pp. 1497-1570, 1983.
7. Dozier, J., "Recent research in hydrology," *Reviews in Geophysics*, Vol. 25, No. 2, pp. 153-161, 1987.
8. Schmugge, T., "Remote sensing applications in hydrology," *Reviews in Geophysics*, Vol. 25, No. 2, pp. 148-152, 1987.
9. Lau, K. M., and P. H. Chan, "Aspects of the 40-50 day oscillation during the northern winter as inferred from outgoing longwave radiation," *Monthly Weather Review*, Vol. 113, pp. 1889-1909, 1985.
10. Lau, K.M. and P. H. Chan, "Short-term variability and atmospheric teleconnections from satellite observed outgoing longwave radiation, Part I: Simultaneous relationships," *Journal of the Atmospheric Sciences*, Vol. 40, pp. 2735-2750, 1983.
11. Butler, D.M., R. J. Gurney, and T. L. Miller, "Eos: The Earth observing system and polar platforms," *EOS, Transactions of the American Geophysical Union*, Vol. 68, no. 45, pp. 1589, 1987.
12. Topographic Science Working Group, "Topographic Science Working Group Report to the Land Processes Branch," Earth Science and Applications Division, NASA Headquarters, Lunar and Planetary Institute, Houston, TX, 64 pp., 1988.
13. Simpson, J., R. F. Adler, and G. R. North, "A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite," *Bulletin of the American Meteorological Society*, Vol. 69, No. 3, pp. 278-295, 1988.
14. JSC Study Group for GEWEX, "Concept of the Global Energy and Water Experiment," Report of the JSC Study Group on GEWEX, Montreal, Canada, June 8-12, 1987, and Pasadena, USA, January 5-9, 1988; WMO/TD-No. 215, WCRP-5, 126 pp., 1988.

IGARSS 89 Abstract

Remote Sensing and Hydrology - The Commercial Prospects

Much is made of the value of remote sensing techniques to geological exploration and a great deal of the remote sensing research effort is centred upon agricultural crop identification and crop-yield prediction. Hydrology has a relationship to both elements. Is there, therefore, a recognition of the value of remote sensing to hydrology and is there a commercial potential?

First, it is necessary to look at the current satellite programme in order to identify what data are available, what forms they take and what are the future trends in instrumentation. These are the "tools" available to the hydrologist.

What are the problems that the hydrologists want to solve; where can remote sensing provide a commercial benefit as opposed to where does the research interest lie?

Given the tools and the requirements, is there a genuine commercial potential? Indeed, how do you measure commercial potential: in terms of revenue to a specialist service or value added industry, or in terms of savings and benefits to the user, or both? Is it possible to estimate the market potential?

It is believed that a commercial potential does indeed exist and that remote sensing offers the hydrologist an invaluable tool now for much of their work and could be essential in the future. If this is true, is the hydrological community doing enough to make sure its voice is heard and its requirements met in future satellite development?

In order to assess this situation, it is necessary to look at some known projects which have provided a commercial benefit to hydrology. This paper will provide selected case studies to demonstrate such projects

J.W. Trevett, President ICRSDT
Associate Director
Hunting Technical Services Limited
Thamesfield House
Boundary Way
Hemel Hempstead
Hertfordshire HP2 7SR US
England

Tel: (442) 23 18 00
Fax: (442) 21 98 88

COST-EFFECTIVENESS OF LANDSAT TM CLASSIFICATION BY OPERATIONS STAFF

S.-L.R. KONRAD & R.S. REMPEL

Ducks Unlimited Canada
1190 Waverley Street
Winnipeg, Manitoba, Canada
R3T 2E2
(204) 477-1760, FAX (204) 452-7560

ABSTRACT

In 1985 NASA completed its feasibility study for Ducks Unlimited (DU) on the use of Landsat thematic mapper (TM) data for mapping wetlands. With the launch of Landsat 5 ensuring a source of TM data for several years, DU began utilizing TM data to inventory wetlands in the Canadian prairie pothole study region (650,000 km² or 250,000 mi²). In completing this inventory and signing the North American Waterfowl Management Plan (NAWMP) in 1988, focus is now toward meeting upland habitat objectives. The approximately 250 spectral classes generated by the unsupervised maximum-likelihood classifier for each scene are already in place for creation of an upland inventory.

To provide data in the time frame required by the NAWMP planners, Ducks Unlimited is decentralizing its image processing capabilities. Micro-computer work stations will be established in priority areas, where operations biological staff can apply their expertise and knowledge of the habitat when classifying scenes. The purchase and implementation of these stations is cost-effective when compared to broad scale acquisition of colour infrared photography at 1:10,000, 1:20,000 or 1:50,000 scales.

The economic values of decentralizing DU's image processing operation are five-fold: 1) ground-truth knowledge and materials are more readily available; hence image classification is very reliable, 2) data processing is more efficient, 3) user needs can be identified more readily, 4) graphics display is available to users during habitat evaluation and planning, and 5) greater flexibility exists for cost-sharing cooperative agreements with other resource agencies, thereby maximizing resource dollars.

DIGITAL IMAGE PROCESSING OF MULTITEMPORAL LANDSAT DATA AND ITS APPLICATIONS IN GROUND WATER EXPLORATION

C.S.S. Reddy, D.J. Campagna and D.W. Levandowski

Department of Earth and Atmospheric Sciences
Purdue University
West Lafayette, IN 47907, USA

ABSTRACT

This research is concerned with developing a procedure for ground water exploration based on digital image processing of multitemporal Landsat data. A test site located south of Tucson, Arizona, was selected to develop the methodology. Three MSS (LANDSAT 1) and a TM (LANDSAT 5) subscenes representing different seasons were geometrically rectified, coregistered and corrected for sun zenith angle and changing illumination effects between dates. The major thrust of the research was directed towards vegetation and drainage texture analysis. Vegetation mapping was based on multitemporal analysis using orthogonal image transformation techniques. Greenness indices of three seasonal data sets were calculated using the Gram-Schmidt orthogonal process and combined to form a three band multitemporal vegetation image. Principal component analysis of this image led to isolation of "dynamic" (crops, etc.) versus static (phreatophytes) vegetation types. A vegetation anomaly map was obtained by classifying these data types. Landsat data (TM bands 4, 5, and 7) were transformed to produce a texture image which is useful in detecting and enhancing drainage texture patterns. The texture image was classified into five distinct hydrologic units, based on their relative permeability values. These results, supplemented with other study elements, such as lineaments and geomorphological maps, obtained from digitally enhanced Landsat data were integrated in a GIS based analysis. A final ground water potential map was obtained by combining "weighted" input GIS files.

Keywords: Ground water, LANDSAT, Digital processing, multitemporal.

INTRODUCTION

In recent years, there has been an increasing demand for ground water supplies because of socio-economic and industrial developments in most of the Third World countries. In this decade, many parts of the world have experienced severe droughts which have resulted in the further depletion of existing ground water sources. This has given rise to a need for development of time and cost effective techniques for the exploration, judicious exploita-

tion, and proper management of ground water.

In the last two decades, there have been several investigations into the application of remotely sensed data for ground water exploration. However, much of the previous work has been typically concerned with the visual interpretation of various types of remote sensor data. For example, aerial photographs are routinely used in preliminary ground water inventories.

Recently several studies have concluded that satellite imagery also can provide valuable information which can be used as a "regional guide" for ground water prospecting, based on visual analysis of landforms, geology, structure, vegetation, drainage density, etc. However, visual interpretation procedures are often subjective, time consuming and cumbersome.

The objective of this research was to develop a procedure for ground water exploration based on digital processing of multispectral Landsat data which paralleled the techniques developed for visual analysis. A test site located south of Tucson, Arizona (Figure 1) was selected to develop the methodology. It is planned to apply the technique to a study of an area in the Mandsaur district of Madhya Pradesh, India. Only the results obtained for the Tucson test site are reported at this time.

Three MSS (Landsat 1) and a TM (Landsat 5) subscenes of the Tucson area, Arizona, representing different seasons were used:

| Path | Row | Date | Scene ID |
|------|-----|---------|---------------|
| 38 | 38 | 2/8/75 | 81930171205G0 |
| 38 | 38 | 4/13/73 | 81264172835G0 |
| 38 | 38 | 8/22/72 | 81030172715G0 |
| 36 | 38 | 12/7/85 | Y5064617261X0 |

PREVIOUS TECHNIQUES

Previous investigations using visual analyses techniques on aerial photographs and Landsat images have noted the surface features of particular importance in ground water exploration are vegetation, drainage density, and lineaments (fracture patterns).

Vegetation

In a comprehensive investigation on plants as indicators of ground water, Meinzer (1927) cited numerous observations on the relation between the depth to the water table and the occurrence of certain plant species in arid regions. He concluded that this relationship can be a useful key for ground water prospecting.

Taranik et al. (1976) outlined a visual interpretation procedure for delineating anomalous vegetation patterns related to ground water occurrence. They described three types of vegetation of particular significance to ground water studies, viz., 1) riparian plants which grow along or in close proximity to river banks, 2) phreatophytes - vegetation which is dependent on ground water supply by tapping the water table through its root system (Meinzer, 1927), and 3) xerophytes - plant species which are able to survive on very small and ephemeral supply of water.

Drainage Density

Drainage texture (in arid regions) is considered to be a good indicator of relative permeability of near surface rocks. Previous studies have demonstrated the applicability of drainage density analysis in inferring lithology, degree of structural control, grain size, relative permeability, compaction, etc. (Salomonson et al., 1983).

Carlston (1963), based on a study of 15 stream basins in the eastern United States, showed that base flow (ground water discharge) is inversely proportional to drainage density. From a mathematical model developed earlier by C.E. Jacob, he derived an equation to show that transmissibility is related to drainage density, to ground water recharge, and to the water table at the water divide.

Lineaments

Lineament analysis is probably the most widely studied and commonly applied remote sensing technique for ground water exploration. Lattman and Parizek (1964) have established the relationship between fracture traces and ground water occurrence based on air photo studies. They have reported high yields for wells drilled on lineaments than for wells located away from lineaments. This work was later followed by many other researchers who concluded that lineament analysis can be a good exploration technique for ground water prospecting.

The determination of these landscape features may be achieved by a variety of procedures. In order to subject as much image data as possible to the scrutiny of the interpreter, separate overlays are prepared for drainage, landforms and lineaments, and vegetation. This approach (Taranik et al., 1976) ensures that most image features are examined and classified to relatively high level of confidence. Eventually the overlays are combined or integrated with ancillary information (geologic and/or soil maps). This analysis procedure results in defining regional trends and relationships and thus establishes a model for exploration or determining drilling sites.

IMAGE ANALYSIS

Preprocessing and Image Registration

The data for the three MSS and TM subscenes representing different seasons were geometrically rectified and co-registered to facilitate temporal analysis.

In addition to normal bad line corrections and rescaling, a correction for sun zenith angle effects and associated effects of changing illumination between dates was applied using a simple cosine ratio function, as suggested by Carr and Schowengerdt (1983), and Ezra et al., (1984). Each data set was separately multiplied by the cosine ratio functions on a pixel by pixel basis. All of the scenes were normalized to a reference zenith angle ($\cos y$) of 39°.

The four Landsat images were geometrically rectified and geo-referenced according to the UTM coordinates. Approximately 20 ground control points (GCP) were obtained from USGS 15' maps. The nearest neighbor algorithm was used to resample the data into 80 x 80 m pixels for the MSS data, and 60 x 60 m pixels for the TM data set. An image-to-image registration was employed using the control points selected from each scene. The April image was chosen as a reference image, to which other data sets were co-registered. Error analysis of the GCP's indicated the accuracy to be within 1/2 pixel space.

Vegetation Analysis

Based on the concept that "static" vegetation (i.e. remains constant through all seasons) such as phreatophytic and/or riparian indicates an availability of ground water whereas "dynamic" vegetation (increases during certain seasons) such as crops, orchards, etc. indicates a dependence on rainfall or seasonal runoff, a procedure was developed to separate and map these types of vegetation. An empirically based image transform, similar to the "Tasseled cap" of Kauth and Thomas (1976) was used to perform the multitemporal vegetation analysis. This transformation rotates and scales the axes of the four dimensional space. The rotated axes of the new coordinate system are termed as brightness (accounts for 98% of soil spectra), greenness (green vegetation), yellowness (yellowing of senescent vegetation), and nonesuch (related to atmospheric conditions).

Site specific coefficients for brightness and greenness were calculated using the Gram-Schmidt orthogonal process (Jackson, 1983). Each set of data was transformed using these coefficients to obtain a brightness and a greenness band. Greenness indices from each of the three data sets were then combined to form a three band multitemporal vegetation image of the study area.

Visual interpretation of this image showed that some areas, especially in the central and northern part of the test site, have nearly the same amount of greenness in all seasons and could represent a possible location of a spring or natural seepage. Intermediate hue values represented phreatophytes or other natural woods. Riparian vegetation was delineated by virtue of its close association with river banks and flood plains. However, crops, orchards, etc., displayed one dominant color

depending on the season and the type of crop present at that time of the year.

Principal component analysis of the multitemporal image was used to decorrelate and isolate these vegetation features into new axes based on their variance values. The "dynamic" vegetation types, such as crops, orchards, etc., define high variance within multitemporal space, whereas phreatophytes/riparian vegetation, "static" vegetation types, are associated with low variance.

Principal component images of the area showed that although PC1 highlighted the overall vegetation patterns of the area, confusion between dynamic versus static vegetation types still existed. PC2 and PC3 isolated most of the riparian and phreatophytes vegetation which are considered to be ground water dependent. This image was then classified into four vegetation classes, using a moving means clustering algorithm. Figure 2 shows the final vegetation anomaly map of the study area.

Drainage Texture

Drainage texture is used to indicate potential areas of ground water permeability and storage. Maximum infiltration occurs on coarse-textured drainage areas whereas fine textures are usually considered to indicate areas of poorer infiltration. Areas having medium- and coarse-textured drainage have streams which are more deeply incised than areas with fine-textured drainage and are more enhanced by low sun angle shadowing on Landsat images. The drainage patterns are distinguishable by different illumination effects (sun-slope geometrics). Coarse-textured drainage channels are widely spaced, have steep sided slopes and gradients. On Landsat images the channels appear dark because of shadows from the side slopes. Medium-textured drainages are less widely spaced, shadows are narrow and light toned in the image. Thus drainage density is proportional to the variance of illumination within a defined cell. A variance analysis of an image of TM 4, 5, and 7 was performed using a 3 x 3 pixel window and normalizing the output file. The data was then classified into five distinct hydrologic units based on their relative permeability values (Figure 3).

Lineament Analysis

Digitally enhanced Landsat TM and MSS data were used to map lineaments. A high pass filter with a 5 x 5 kernel size was used to enhance linear features in the TM data. This image was displayed on the high resolution color monitor screen, and interpreted lineaments were then digitized directly from the screen. The analysis was also supplemented by a regional scale geological map, color, and B&W hard copy prints of TM and MSS data at 1:1,000,000 scale.

ANCILLARY DATA

Geomorphological Map

A geomorphic map of the area was prepared by visually interpreting hard-copy prints of digitally enhanced TM images and using results from the drainage texture analysis. A map displaying nine geomorphic units was prepared and digitized into the GIS file.

Bedrock Outcrop and Basin Fill Thickness Maps

Published maps of the geology (Wilson et al., 1960) and alluvium thickness (Cooley, 1973) was digitized and entered into the database of the GIS File. Both maps were geo-referenced to UTM coordinates and a pixel size of 80m x 80m was used to facilitate a "map overlay" based analysis.

INTEGRATED ANALYSIS

The GIS files compiled from previous studies along with Landsat derived files were analyzed in an integrated model based approach. In order to isolate and target areas of interest based on the concept of "convergence of evidences". The model functions as a simple additive scheme where areas of interest are highlighted and those areas that are not significant to this work are masked. The analyses reported herein, were performed using a raster based GIS software provided with the ERDAS image processing system. A brief description of the procedure is outlined below:

1/selection of the criteria for data analyses. The study elements were integrated by combining two GIS files at a time.

2/recoding of the class values of an input GIS file to mask out certain data values which have lesser significance.

3/proximity analysis of lineaments to assign an area of influence. The lineaments occurring in basin areas were assigned a field of influence of 240m (3 pixels) in length.

4/each class in the input file was multiplied by weighting factors. The actual value of the weighting factor depends on the relative importance of the data being analyzed. The more important the criteria, the higher the weighting factor. Weighting factors of 3, 2, 2, and 1 was assigned to vegetation, geomorphology, lineaments, and drainage texture respectively.

5/normalization of output files was performed by dividing by the sum of weight factors, thereby reducing the number of classes in the output file.

6/overlaying bedrock mask obtained from alluvial thickness and geology maps, onto the final water potential map.

The final ground water target map obtained by applying the procedure described above is shown in Figure 4. The results correspond closely to the results presented by Taranik et al. (1976) for a visual analysis of the same area.

CONCLUSIONS

Some of the advantages offered by the methods reported in this work are: 1) they are based on a model approach which incorporates temporal and a variety of other data types, 2) the proposed procedure allows the user to quantitatively map potential areas of interest based on various criteria describe this work, 3) it utilizes data reduction technique which are an important consideration in handling multitemporal data, and 4) it is an objective oriented procedure and makes repetitive analyses possible.

REFERENCES

Carlston, C.W., "Physiographic and hydraulic studies of rivers, drainage density and streamflow", USGS Prof. Paper 422-C, pp 1-8, 1963.

Carr, J.R., Schowengerdt, R.A., "Signature extension versus retraining for multispectral classification of surface mines in arid regions", Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 8, pp 1193-1199, 1983.

Cooley, M.E., "Map showing distribution and estimated thickness of alluvial deposits in the Tucson area, Arizona", U.S. Geological Survey, Misc. Investigations Series, Map I-844c, 1973.

Ezra, C.E., Tinney, L.R., and Jackson, R.D., "Considerations for implementing vegetation indices for agricultural applications", Technical Papers, 50th Annual Meeting of the American Society of Photogrammetry, pp 526-536, 1984.

Jackson, R.D., "Spectral indices in n-space", Remote Sensing of Environment, Vol. 13, pp 409-421, 1986.

Kauth, R.J., and Thomas, G.S., "The tasselled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat",

Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN, pp 41-51, 1976.

Lattman, L.H., and Parizek, R.R., "Relationship between fracture traces and the occurrence of ground water in carbonate rocks", Journal of Hydrogeology, Vol. 2, No. 2, pp 73-91, 1964.

Meinzer, O.E., "Plants as indicators of ground water", USGS Water Supply Paper No. 577, pp 1-99, 1927.

Salomonson, V.V., (author-editor), "Water Resources Assessment", in Manual of Remote Sensing. Colwell, R.N., editor-in-chief, American Society of Photogrammetry, Vol. 2, p 1502, 1983.

Taranik, J.V., Moore, G.K., and Sheehan, C.A., "Targeting groundwater exploration in south-central Arizona using Landsat satellite imagery", USGS Workshop exercise, EROS Data Center, Sioux Falls, SD, pp 1-34, 1976.

Wilson, E.D., Moore, R.T., and O'Haire, R.T., "Geologic map of Pima and Santa Cruz counties, Arizona", Published by Arizona Bureau of Mines, University of Arizona, Tucson, AZ, 1960.

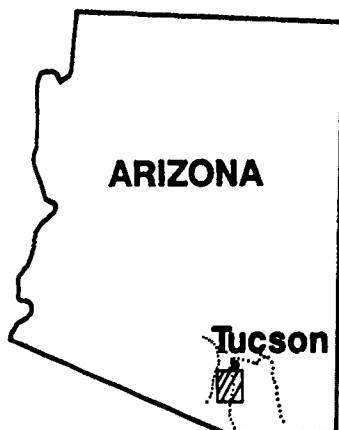


Figure 1. Location map of test site.



Figure 2. Vegetation Anomaly map.

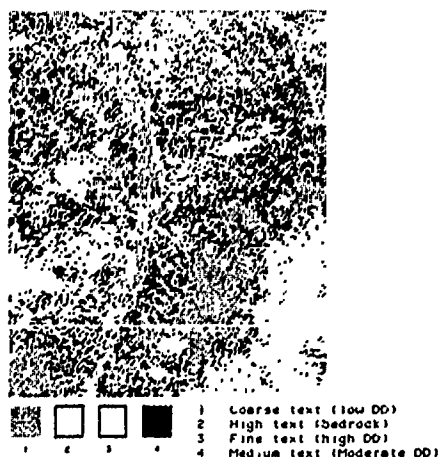


Figure 3. Drainage (Density) Texture map.

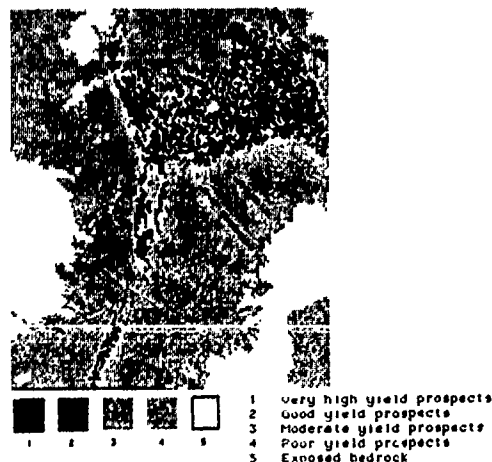


Figure 4. Ground Water Potential map.

COMBINING REMOTE SENSING AND CONVENTIONAL DATA FOR ASSESSING WATER RESOURCES IN NORTHERN THAILAND

C. Saowapon¹, S.I. Solomon², and E.D. Soulis²

¹Department of Civil Engineering, Chiang Mai University, Chiang Mai, Thailand 50002.

²Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1.

ABSTRACT

This paper examines the usefulness of remotely sensed data, in particular data from Landsat, to estimate monthly flows at ungauged sites. A deterministic, lumped-input, lumped-parameter hydrologic model is used to estimate flow of ungauged river basins. Two techniques that relate the model parameters to basin characteristics required by this model are assessed. First, a multiple regression analysis using reflectance values of Landsat MSS as independent variables is used to transfer the model parameters to ungauged sites. A second technique is a vegetation index method that uses a visual segmentation of Bands 5 and 7 images into grassland and woodland. Model parameters are estimated as weighted averages of values derived elsewhere. The results in both cases are comparable with those obtained using the Thiessen polygon technique for parameter interpolation.

KEY WORDS: Remote sensing, hydrologic models, Thailand.

INTRODUCTION

The need for long-range planning of water resources has become more evident in recent years with population growth and increased development and utilization of the world's resources. Since manpower and funding are limited, priorities for resource development and utilization must be established through systematic planning. Good quality hydrological data are necessary for the efficient planning and design of all water management projects. Unfortunately, in many areas hydrological measuring stations are too few, and in some areas nonexistent. Herschy et al. (1985) reported that there were some 130,000 rain gauges in the global network, most of these in developed countries. At the present time elsewhere most traditional means of data gathering are provided by low technology and high labour costs and they are inadequate even for day to day management of water resources. Therefore, hydrologists are assiduously seeking new ways of augmenting their conventional data supplies. Remote sensing is being explored as one possible answer to such data acquisition problems, not the least because of its perceived facility for providing greatly improved space and time coverage of key hydrological variables.

The hypothesis of this research is that the reflectance characteristics of the drainage basins obtained from Landsat MSS data correlate with their physical characteristics, e.g. vegetation, soil type, etc., and that these in turn can be related to model parameters that are physically meaningful and therefore vary in accordance with the basin characteristics. When a hydrologic model which uses such parameters is applied, the MSS data should improve the estimation of the model parameters for ungauged basins.

THE STUDY AREA AND DATA AVAILABLE

The study area is located in the upper part of the Ping River basin, which extends in two provinces of Chiang Mai and Lamphun in northern Thailand. The total area of the basin being studied is 1,274,500 hectares and is subdivided into 17 sub-basins. About two-thirds of the basin area is mountainous and covered mainly by mixed deciduous forest. Evergreen, hill evergreen, pine, and deciduous dipterocarp forest are also found in this area (Klankamsorn and Charupatt, 1981). The rest is a rich flood plain area.

To avoid flow data affected by diversion, only basins that did not have water diversion were selected for this study. Thus only 17 stations of the 27 Royal Irrigation Department (RID) flow measuring stations were selected. Precipitation data were obtained from 8 stations operated by the Meteorological Department and 6 stations operated by the National Energy Administration (NEA). The temperature data were obtained from 5 NEA's stations. Thus, the number of precipitation and temperature stations in the area is less than the number of flow gauging stations. This is explained by the great interest in using the river water in the area for irrigation purposes. From a hydrologic view point, this area is one of the best gauged areas of Thailand and was selected because it offers potential for split sampling techniques of calibration-validation. However, unfortunately the area is poorly gauged from a meteorological view point. This unusual situation resulted in some difficulties in the use of rainfall-runoff model in this basin (Saowapon, 1987). The study period was January 1971 to December 1980. However, the study concentrated on the 1977 to 1980 interval, because data are available at all stations during this interval, and because it was felt that more recent data are more accurate. Also, river basin

conditions during this period were probably closer to those observed in 1983 and 1985, years for which Landsat reflection data were available.

Two Landsat 4 MSS images -Path 131, Row 47- obtained by the Landsat receiving station in Thailand were used in this study. The dates of the images are April 24, 1983 and March 28, 1985. Both images were obtained on computer compatible tape (CCT) format and as 1:250,000 black and white photographic products in Band 5 and Band 7. Both images, which have less than 5% cloud cover, were obtained during the dry season, since during the wet season the area is normally covered by clouds.

HYDROLOGIC MODEL

The model selected was one which is simple and has been proven to have parameters related to the basin characteristics (Wishart, 1987). The model is essentially a deterministic, lumped-input, lumped-parameter type model. It has four coefficients for estimating interflow, infiltration/runoff, percolation, and evapotranspiration. There are no components for baseflow or channel routing. Therefore, the model sees precipitation as either running off the land and leaving the system or remaining as infiltration. The infiltration water then returns to the channel through interflow or returns to the atmosphere through evapotranspiration or is lost to the system as seepage below the root zone to the phreatic zone.

The model has 4 parameters, which are listed in Table 1 along with their definitions. A basic philosophy of the model is to recognize explicitly the spatial variability of each parameter, that is the value of any parameter depends on the characteristics of the particular drainage basin. Therefore, the model parameters need to be calibrated before the model is applied to any drainage basin. However, the parameters are not considered to vary in time as long as the river basin characteristics remain the same. Change in land-use/land-cover may result in changes in model parameters.

Table 1 Limit value of model parameters

| | DEFINITION | LIMIT VALUES |
|-------|--------------------------------------|--------------|
| SEC | Soil evapotranspiration coefficient | 1 - 12 |
| SMC | Soil moisture capacity (mm) | 0 - 500 |
| SFC | Subsurface flow coefficient (mm) | 0 - 200 |
| UPGWR | Upper limit of ground-water recharge | 0 - 1 |

Rosenbrock's method of hill climbing (Rosenbrock, 1960) was selected as the optimization technique used to find the value of each parameter during the model calibration. The objective function used in parameter calibration measures the standard error of estimated flow. The optimum model parameter corresponds to the minimum objective function.

There are two basic inputs to the model: monthly precipitation and monthly average temperatures. However, during

the calibration of the model's parameters, monthly runoff is also a required input. Monthly runoff and monthly evapotranspiration are the main outputs of the model. The model also provides the value of monthly percolation and soil moisture estimates as additional outputs.

BASIN REFLECTANCE ANALYSIS

MACHINE ANALYSIS

The variation of Landsat spectral signatures from one frame to another is dependent upon a number of factors, such as changes in solar elevation angle, differences in intrinsic reflectance of surface targets, and effects due to optical thickness and composition of the atmosphere. The effect of these atmospheric variabilities needs to be eliminated before using these data for reflectance analysis. In this study, the atmospheric calibration process was completed at the Canada Centre for Remote Sensing (CCRS) in Ottawa.

All 17 watersheds were delineated and registered to the 1983 image. Geometric correction was done to transform the 1985 image co-ordinates into the co-ordinates of the 1983 image by using 30 control points. The watershed boundaries were transferred to the 1985 image from the 1983 image. The mean reflectance of each basin in four different bands was obtained.

MANUAL INTERPRETATION

Four 1:250,000 scale Landsat 4 MSS photographic images in Band 5 and Band 7 were used for manual visual interpretation. The watershed boundaries, for which the co-ordinates had been transformed from UTM to Landsat-based, were traced on transparent films. Using these films, the watershed boundaries were located on each Landsat image.

The reflectance for each image was divided into 5 classes. The brightest class was assigned a reflectance index of 5 and the darkness of 1. Indices were assigned to each element of a square grid superimposed on the two images. Figure 1 plots the relationship between reflectance values obtained from visual interpretation with values obtained from machine interpretation.

In addition, the land in each basin was classified into two types - woodland and grassland- according to their reflectance characteristics. In each classification farmland was combined with grassland. Taking into account the

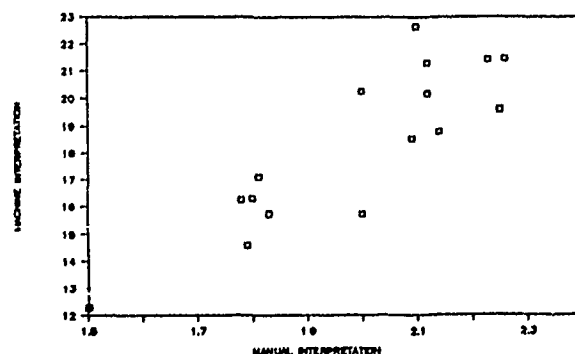


FIGURE 1 Reflectance values of Band 7 (1983 image)

reflectance properties of dense and sparse vegetation, if the area appeared as bright in Band 5 and dark in Band 7 it was classified to grassland, and areas appearing dark in Band 5 and bright in Band 7 as woodland.

RESULTS

MODEL CALIBRATION

The model was calibrated for each basin by using the optimization technique. Figure 2 shows the plot of the measured flows versus estimated flows. The modelled peak flows generally correspond well with those observed.

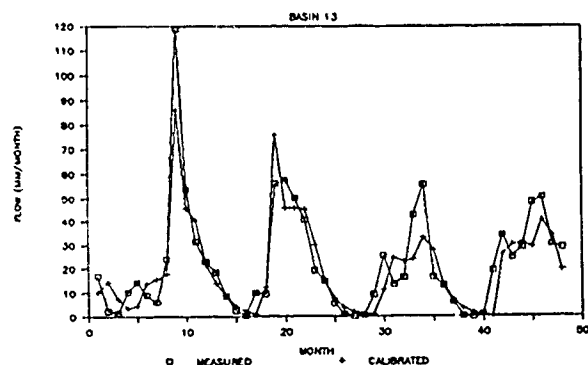


FIGURE 2 Hydrograph of measured flow and calibrated flow

MODEL VALIDATION IN TIME

The model validation process consists of validation in time and space. The validation in time consists of applying the calibrated parameters to a time period other than the one used for calibration. This process was carried out on 7 river basins. The basins used in this phase had 48 months of data available. The first 36 months of data were used to calibrate the model parameters. These parameters were applied to the last 12 months. The standard errors of validation are of the same magnitude as the standard errors of calibration (Saowapon, 1987). This indicates that the model is acceptable for simulation of flows during other periods of time for which only precipitation and temperature data are available.

RELATIONSHIP BETWEEN REFLECTANCE VALUES AND MODEL PARAMETERS-VALIDATION IN SPACE

The main objective of this analysis was to find if a relationship between reflectance values and model parameters can be obtained. In order to check the value of such a relationship for estimating model parameters in ungauged areas, a split sampling technique of calibration-validation was used. From the 17 gauging basins, four sets of samples were used for the calibration and validation process. In the first sample, sample A, 14 basins were used for calibration and 3 basins, which were assumed to be ungauged basins, were used for validation. In the other three samples (samples B, C, and D), 12, 10, and 8 basins were used for calibration and 5, 7, and 9 basins were used for validation respectively.

The model parameters for ungauged basins were calculated by three different techniques. The first approach involved

TABLE 2 Model parameters obtained from basins in Africa (Wishart, 1987)

| TYPE OF LAND-COVER | SEC | SMC (mm) | UFGWR (mm/month) | SI-C |
|--------------------|------|----------|------------------|------|
| Grassland | 10.1 | 52.0 | 58.2 | 0.55 |
| Woodland | 8.0 | 185.0 | 43.6 | 0.34 |
| Rain-forest | 6.3 | 374.0 | 32.4 | 0.22 |

a variety of multiple regression analyses. Model parameters, obtained by calibration, were regressed with Bands 5 and 7 from manual interpretation and from machine interpretation and with Bands 4 and 6 from machine interpretation. The results were used to calculate parameters for ungauged basins. If the value of the calculated parameter was higher than the upper limit of that parameter, the upper limit would be assigned to it. On the other hand, the lower limit would be assigned, if the calculated value was less than the lower limit of that parameter.

The second technique is a vegetation index method that uses the classified grassland and woodland areas as weights to estimate the model parameters as weighted averages of the parameter values as shown in Table 2. The values were developed for Africa by Wishart (1987).

Thirdly, the Thiessen Polygon technique was selected as the conventional technique to interpolate model parameters from the nearby basins to the ungauged basin. The results of this technique were used as the basis to compare the improvement of flow estimation using other techniques.

The arithmetic mean of the error of each technique of model parameters estimation was calculated. The mean of the absolute value of the error (MAVE) and its standard error were also calculated. Table 3 shows the summary of error analysis of alternate techniques for the flow estimation.

The Thiessen Polygon technique provided the lowest standard error when using sample A, C, and D as the calibration sample with the standard error of 111.2 %, 87.1 % and 93.9 % and the MAVE of 62.2 %, 44.4 %, and 53.6 % respectively. On the other hand, the simple landcover-classification technique gave the lowest MAVE for validation of every validation sample. Note that in this case there is no model calibration as the parameter values were obtained from the results of a different study in relation to the estimated vegetation characteristics of each basin. The two machine analysis techniques, i.e. Band 4 and Band 6, and Band 5 and Band 7, had a slight difference in the MAVE. However, the standard error of the combination of Band 4 and Band 6 are greater than the standard error of Band 5 and Band 7. The visual interpretation provided lower MAVE and standard errors than the machine interpretation. For each technique, the MAVE showed an decreasing trend with the increase of the number of calibration basins.

The Thiessen Polygon and vegetation index methods provided total estimated flows in the same direction, i.e. both overestimated or under estimated, only three river basins have results in opposite direction. The total estimated flows for small basins (areas less than 5,000 hectares) are underestimated. Figure 3 plots the estimated flows obtained from both techniques versus the measured flow of basin 13. In general, the estimated flow from both techniques correspond acceptably well. However, the estimated

TABLE 3 Error analysis (in percent) of alternate techniques for flow estimation

| No of basins used for estimation | STATISTICS | SOURCE OF MODEL PARAMETERS ESTIMATION | | | | Land Classification |
|--|----------------|---------------------------------------|--|------------------------|---------------------|---------------------|
| | | (Thiessen Polygon) | Visual Interpretation (Band 5 and Band 7) | Machine Interpretation | | |
| | | | | (Band 5 and Band 7) | (Band 4 and Band 6) | |
| Station A (16) | Abs. error | 51.8 | -49.1 | 26.2 | 36.4 | 24.3 |
| | MAVE | 62.2 | 62.2 | 61.6 | 71.9 | 50.3 |
| | Standard error | 112.2 | 120.1 | 137.2 | 148.0 | 121.0 |
| Station B (12) | Abs. error | 8.1 | 6.8 | 29.5 | 35.8 | 23.2 |
| | MAVE | 45.0 | 50.2 | 46.3 | 52.1 | 43.0 |
| | Standard error | 95.0 | 85.1 | 84.9 | 154.8 | 89.6 |
| Station C (10) | Abs. error | 4.4 | 4.7 | 9.6 | 8.5 | 18.5 |
| | MAVE | 41.4 | 52.1 | 62.2 | 55.1 | 39.2 |
| | Standard error | 37.1 | 98.9 | 107.9 | 110.3 | 95.9 |
| Station D (8) | Abs. error | 9.5 | -7.7 | 19.3 | 15 | 9.0 |
| | MAVE | 55.6 | 166.3 | 58.9 | 61.5 | 49.4 |
| | Standard error | 93.9 | 113.2 | 124.3 | 174.9 | 101.0 |

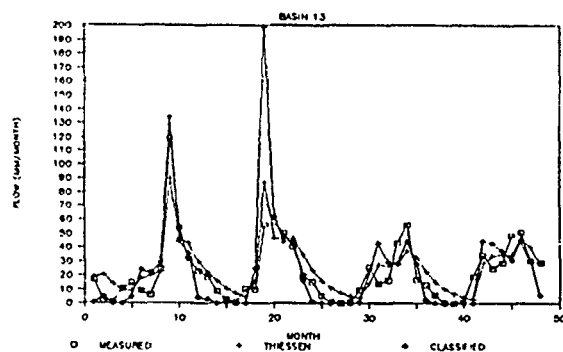


FIGURE 3 Hydrograph obtained from Thiessen Polygon and basin-classification compare to measured hydrograph

flows obtained through basin classifications show zero estimated flow during the dry season followed by rapid flow increases. This result indicates that the subsurface flow coefficient and soil moisture capacity which were obtained from the study area in Africa do not fit well for this area. Further research, using additional information on the soil type and the geology of the area and their relationship to the model parameters, is therefore recommended.

CONCLUSIONS

1. The deterministic, lumped-input, lumped-parameter type model can be used for flow data synthesis at ungauged stations in Northern Thailand.
2. The mean reflectance of the basin from Landsat MSS, from both machine and visual analysis, can be used to estimate the model parameters with an accuracy comparable to that obtainable by interpolation of model parameters using Thiessen polygons.
3. The land-cover classes of woodland and grass land which were classified from Bands 5 and 7 yield acceptable results in estimating model parameters. Model parameters are estimated successfully without calibration using an African relationship and Thai land-cover as input. This is an important result for regions not as heavily gauged as this study area.
4. The remote sensing techniques do not provide significant improvement of flow estimation when compared with the conventional technique of Thiessen polygon.

However, this conclusion is probably valid only for densely gauged areas. The research provided evidence that for ungauged areas the estimation of model parameters from vegetation estimates is effective.

ACKNOWLEDGEMENTS

The first author was supported to conduct this research at the University of Waterloo by the Canadian International Development Agency and their support is gratefully acknowledged. The Remote Sensing Division, National Research Council, Bangkok, provided the Landsat data. The CCRS, Ottawa, provided their CIAS facility for analysis of the Landsat data.

REFERENCES

1. Herschy, R.W., Barrett, E.C., and J.N. Roozkrans, "Satellite remote sensing in Hydrology", Proc. EAR-SeL/ESA Symposium on European Remote Sensing Opportunities: System, Sensors and Applications, Strasbourg, France, pp 47-53, 1985.
2. Klankamsorn and Charupatt, "Study on the changes in forest land-use in Northern, Thailand by using Landsat Imagery", (in Thai) Royal Forest Department, Bangkok, Thailand, 1981.
3. Rosenbrock, H.H., "An automatic method for finding the greatest or least value of a function", Computer Journal, Vol. 3, pp 175-184, 1960.
4. Saowapon, C., "Combining remote sensing and conventional data for assessing water resources in Northern Thailand", MA.Sc. Thesis, Department of Civil Engineering, University of Waterloo, 1987.
5. Wishart, W.D., "A hydrologic model for the estimation of water resources in Africa", MA.Sc. Thesis, Department of Civil Engineering, University of Waterloo, 1987.

USING NOAA DATA FOR HYDROLOGIC MODELING

GEOFF KITE

HYDROMETEOROLOGICAL RESEARCH DIVISION
CANADA CLIMATE CENTRE
NATIONAL HYDROLOGY RESEARCH CENTRE
11 INNOVATION BLVD., SASKATOON, SASK., S7N 3H5, CANADA

ABSTRACT

The paper describes a study to use remotely sensed data, together with conventional ground-based meteorological and hydrological data for watershed modeling. The data are combined in a database written specifically for hydrometeorological modeling on micro-computers.

Two basins in western Canada with significant snowpack are being studied; the Kootenay and the Souris and two distributed models are being developed. Daily visible and near infra-red data from a NOAA satellite are used, both from digital tape and from black and white photographs.

The objectives of the study are to investigate the utility of data from satellite in a hydrometeorological model in different physiographic and climatic areas, to use the model as a test-bed for different physically-based components and to develop a land-phase component for climatic general circulation models.

1. INTRODUCTION

Four major developments are occurring in hydrological modeling:

- (i) Development of models is increasingly towards the use of distributed variables and parameters.
- (ii) Models are becoming more physically based.
- (iii) New sources of data are being exploited; particularly remotely-sensed data.
- (iv) Hydrology, climatology and oceanography are being coordinated for global circulation models to study the present climate as well as possible climate fluctuations.

The study described uses existing hydrologic models, developing them towards a more physics-based state, and towards the condition where satellite data is used as principal data input. The eventual aim is to develop one of the models for use as a land-phase component for use in GCMs.

2. MODELS

2.1 The need for a Hydrological Component within GCM's

Mathematical models of the global circulation (GCM's) are amongst the most complex and computation-intensive models in the natural sciences. Through both national research initiatives and international programmes such as GARP (Global Atmospheric Research Programme) and WCRP (World Climate Research Programme) major progress has been made in understanding atmospheric processes.

Similarly, on-going programmes such as TOGA (Tropical Ocean and Global Atmosphere) and WOCE (World Ocean Circulation Experiment) are expected to produce significant improvements in understanding of the ocean dynamics and their interaction with the atmosphere.

These developments have led to the construction of many global circulation models (GCM's) and numerical weather prediction (NWP) models. Most GCM's contain a land-phase component, but this is probably the least understood part of the water cycle and is, perhaps, the weakest link in the GCM's.

Although the total annual transfer of water from land to sea (streamflow), estimated (NRC, 1986) at $36 \times 10^3 \text{ km}^3$, is small compared to the total annual precipitation (over land and sea) of around 500×10^3

km^3 , the geographical distribution of land surface parameters is of particular importance to GCM's. The existing hydrological components usually account only for changes in soil moisture and snow cover; improvements in GCM's could be made in the treatment of vegetative cover, soil water capacity, water movement in soils, the groundwater system and in surface runoff.

Adding a more detailed hydrometeorological component to GCM's would also lead to better integrated models of the global climate system since the generated runoff data could be compared to recorded streamflow data as a check on the performance of the GCM.

2.2. Hydrologic Models

A hydrologic model is an attempt to represent in understandable form the complex system controlling the land-phase of the water cycle. In essence, hydrology, as a science, is the development of such models. Many models have been developed for many different purposes, and many classification schemes have been proposed (e.g. Clarke, 1973; Fleming, 1979).

Simply, hydrologic models can be categorised in two ways; by the degree of understanding of the catchment physics involved and by the treatment of the basin physiography (Figure 1). A stochastic model assumes no knowledge of the physics of the basin, a parametric model assumes some limited knowledge and a fully physical model uses the known laws of physics to relate basin input to basin output. Lumped models use basin-average data, distributed models use the physiographic knowledge of the basin as important information.

Stochastic models and parametric models are generally lumped while physical models must be distributed by nature.

2.3. Choice of a Hydrologic Component for a GCM

The purpose of a general circulation model is to simulate the climate of the atmosphere from first principles (Boe, et al, 1984). A hydrologic component must therefore be as physically-based as possible.

GCMs operate with grid squares in the order of $200,000 \text{ km}^2$ compared to the 10 km^2 to $10,000 \text{ km}^2$ scale of hydrologic models, so the interaction between the two models must be averaged. Similarly the time scales of the two models are different, introducing another averaging.

Physical models need immense numbers of data and the only practical source of these data is remote sensing.

Based on previous studies of grid square models (Solomon et al, 1968), physically based models (Charbonneau et al, 1977) and on the use of remotely-sensed data (Robinson & Associates, 1986) it was decided to use the distributed semi-physical development of CEGUEN (known as the Hydrotel model, Fortin et al, 1988) as the main model, together with a development of a much simpler model (Kite, 1978) as a check.

2.4 Hydrotel Model

The Hydrotel model was developed by INRS-Eau at the University of Quebec (Fortin et al, 1988) for use on microcomputers. There are seven modules, written in the 'C' language, for data input, physiography, precipitation, evapotranspiration, ground-water hydrology, optimisation and output.

The model operates on a menu-submenu system, the user selecting the required operations of the model and specifying the necessary parameters. A basin 'mask' defines the arrangement of grid squares over the basin and further files define the channel reaches and the channel nodes. During a simulation run, the user can display the status of the variables and parameters for any grid square, channel reach or channel node at succes-

sive time steps. After a simulation run, temperatures, rainfall, snowmelt, outflow, surface runoff, and streamflow can all be stored for each grid square for each time interval.

2.5 SLURP Model

The SLURP model is a simple lumped basin model used as a check on the data and on the modeling results; a simple model can often detect errors in recorded data. This model has previously been used in a comparison with the SSARR model (Kite, 1974). The model uses basin-averaged precipitation and mean daily temperature from the database described in the next section. Snowmelt is calculated by the degree-day method and evaporation/evapotranspiration is calculated using the complementary relationship (Morton, 1983).

3. DATA

3.1 Ground-based data

Table 1 summarises the types of data used in the hydrologic models. Ground-based data includes climatic data such as daily maximum and minimum temperatures, daily precipitation and daily snow on the ground as well as hydrometric data, physiographic and land-use data. Provision had therefore to be made for both point and distributed data.

3.2 Satellite Data

The types of data which satellites can provide for hydrological models include snow cover (Allen & Mosher, 1986), snow water equivalent (Goodison et al., 1986), cloud cover and precipitation (Moses & Barrett, 1986; Tsanis & Isaac, 1985) and evapotranspiration (Hatfield et al., 1983).

At this stage of the study, the only data needed from satellite are the percentages of each grid square within the two basins which are covered by snow and/or cloud on each day. The practical choice of satellites which could provide these data include (Table 2) the polar-orbiting Landsat series using multi-spectral scanner (MSS) or thematic mapper (TM) data, the sun-synchronous, SPOT satellite, the NOAA series of polar-orbiting satellites and the GOES series of geostationary satellites.

SPOT has the highest resolution (10-20 m.) but was rejected on cost grounds. Landsat offered next best resolution (30 m. for TM and 80 m. for MSS) but, again, the data were too expensive to use on a daily basis. The data from GOES, although available at higher than required frequency, were not used because of the low resolution and the high level of distortion at the latitudes of the study basins.

The study uses daily visible and near infra-red data (bands 1 and 2) from a NOAA satellite for cloud and snow information as well as one Landsat image per basin for land classification.

3.3 Database

After a review of commercial database packages it was decided to write a database for a microcomputer in Fortran using a full-screen data entry system. The chief advantage of this system over proprietary database languages is that the data are in files which are directly accessible by programs written in Fortran or C. Using a proprietary database languages might have meant keeping two sets of parallel files, possibly leading to confusion.

The database starts with a main menu with options of updating or printing data, displaying data distributed over the grid squares of the basin, calculating the distributed data or adding a new basin to the database. Distributed data include precipitation, temperature, snowdepth, elevation, slope, aspect, and land-use types.

If the user is adding or modifying data then the appropriate form will be screened (Figures 2) and, if the data exist they will be located using direct access files and will be written to screen for viewing or update. Backing up each form is an extensive help, error and validation system to ensure data quality. After filling or correcting the form the user can scroll forward or backward a month at a time through that station's record or can choose another station or return to the main menu.

4. BASINS

Three criteria were used in the selection of basins to study:

- i) They should be in different physiographic and climatic zones within Western Canada.
- ii) They should be large enough to enable use of NOAA images.
- iii) They should have routine hydrometeorologic data available.

The basins selected are the Kootenay and the Souris. The Kootenay basin is situated in the Rocky Mountains in south-eastern British Columbia while the Souris covers parts of south-eastern Saskatchewan, south-western Manitoba and northern North Dakota. Figure 3 shows 3-dimensional diagrams of the two basins.

The Souris basin has one complication not present in the Kootenay; non-contributing area. In a prairie watershed, not all runoff reaches the main stream; because of the topography, some runoff is caught in depressions or sloughs with internal drainage and never reaches the main

stream. The effective real drainage area varies with antecedent conditions and with precipitation in a complex manner. In order to deal with this in a model, the effective drainage area is defined as that area of a basin which might be expected to contribute runoff during a year having runoff with a return period of two years. The effective drainage areas for Alberta, Saskatchewan and Manitoba have been mapped (PFRA, 1983) and these maps have been used to delineate effective drainage area for each grid square in the Souris basin.

5. DATA ANALYSIS

5.1. Satellite Data on Magnetic Tape

NOAA satellite data for the two basins are received from the Atmospheric Environment Service receiving station in Edmonton, Alberta, on magnetic tape. The data are transferred from tape to a 70 Mb hard disk on a 386 microcomputer using a 1600/6250 bpi 9-track tape drive. The hard disk on the micro was partitioned into a C: drive of 20 Mb for programs and a 50 Mb D: drive for image data. The transfer program allows the user to select a variable sized window of data from the full image on tape.

Once the data are on the hard drive, a commercial software package is used to display and analyse the image. A particularly clear image for each basin was geographically corrected to the appropriate grid square, and all subsequent images are registered to these base images. The cloud-covered and snow-covered areas within the basins are identified using a supervised classification technique and the percent of each grid square covered by cloud, snow, both or neither is calculated (Kite, 1988). At any stage in the image analysis process 35mm slides and 8.5 x 10.8 cm. colour prints can be obtained using an on-line image capture device. Figure 4 shows a flowchart of the image analysis procedure.

A series of computer programs were written to automate these procedures but, still, a high degree of manual intervention is needed for navigation of the image and for classification.

5.2. Satellite Data as Photographic Images

Black and white photographs of NOAA visible and IR images (Figure 5) are used as a quicker and easier method of obtaining cloud and snow cover for the basins for interpolation between the images from digital tape. The method used for this analysis has been described by Johnstone and Ishida (1984) and was originally used for flood forecasting on the St John River, New Brunswick. The equipment consists of a dedicated microcomputer, a video camera and copy stand and a density slicing unit. Figure 6 shows a flow chart of the image analysis procedure.

The image is navigated to a pre-digitised basin outline containing geographical reference points by moving the photograph and zooming the video camera. The image is video digitised and the resulting digital data are analysed using density slicing to select suitable thresholds for cloud and snow cover. Finally, a program calculates the percent of each grid square covered by cloud, snow, both or neither. Usually the visible image is used for snow on the ground and the IR is used to measure cloud.

Again, a high degree of manual intervention is needed, but the average time needed to analyse an image is only 20 minutes compared to 2 hours using the system from magnetic tape.

Table 3 shows some statistics for the NOAA black and white images used. It can be seen that for the Kootenay basin, over 90% of the images expected were received and about 89% of those expected were both received and usable. Of the usable images received, only 36% were for non-100% cloud days (that is, days for which snow-cover could be measured). Comparable statistics for the Souris basin are 93%, 91% and 48%. Reasons for images being unusable include poor alignment (missing part or all of the basin) and streaky images resulting from software problems at the satellite reception station. The software problems are associated with a major changeover in reception systems.

Such statistics would be useful for developing a research model such as this into a practical model for water resources engineering.

5.3 Combining Satellite and Ground-based Data

The database contains an option for combining satellite and ground-based data. In this option temperature data are distributed to each grid square using data from the three closest climate stations adjusted for elevation and weighted according to inverse distance. Precipitation is distributed similarly and is then adjusted using the satellite cloudcover information. Snowcover distribution uses the additional information from satellite snowcover data and from snowcourse water equivalent data. Temperature and precipitation lapse rates are variables and the user can also manually adjust any of the distributed data.

6. DISCUSSION

The U.S. National Research Council (1986), in its proposal for an International Geosphere-Biosphere Program, concluded that,

'Major efforts should be made to develop conceptual models to better understand why changes are taking place in the hydrologic cycle and to integrate these models with global-scale atmospheric and oceanic models for study of climate

change. Analytical work is needed to develop hydrologic models that successfully simulate such factors as streamflow.

Such work will not only improve the GCM's but may also improve understanding of the physical processes of the watershed, since, according to Klemes (1982):

'It seems very likely that hydrology will jump ahead after it's links with processes at the planetary level are better established.....This belief stems from an observation that a successful solution of a problem is more likely if it is approached from two opposite directions. In hydrology, the "other" direction is "downwards" from global concepts. It follows that the remote sensing techniques, oriented to areal rather than point events, may have an inherent potential for fertilization of hydrological thinking'.

It has also been argued (Bevan & Hornberger, 1981) that as precipitation is the most variable meteorological element, errors in estimating precipitation are likely the limiting factor in runoff simulation and that conventional means of estimating precipitation are inadequate. Areal estimates from satellite offer a means of overcoming these limitations.

The first steps in this project have been to design a database to combine remotely-sensed data and ground-based data. Data for two basins for an 18-month period have been collected and entered in the database. Two hydrologic models have been obtained, one distributed and one lumped. The models will be applied to the two basins to estimate streamflow and the utility of the remotely-sensed data will be assessed. Later stages of the project will develop the physically-based components of the distributed model and adapt it as a land-phase component for GCM's.

7. REFERENCES

- Allen, M.W. & F.R. Moshier, "Monitoring Snowpack Conditions with an Interactive Geostationary Satellite System", Proc. 54th Annual Western Snow Conference, Phoenix, Arizona, 51-60, 1986.
- Bevan, K.J. & G.M. Hornberger, "The Effects of Spatial Variability in Precipitation on Streamflow", Technical Report No. 6, State Water Survey Division, University of Illinois, Champaign, Illinois, 1981.
- Boer, G.J., N.A. McFarlane, R. Laprise & J.D. Henderson, "The Climatology of the Canadian Climate Centre General Circulation Model as Obtained from a Five-Year Simulation", Atmosphere-Ocean, Vol. 22, no. 4, pp 430-473, 1984.
- Carroll, T. & M.W. Allen, "Airborne Gamma Radiation Snow Water Equivalent and Soil Moisture Measurements and Satellite Areal Extent of Snow Cover Measurements, a User Guide, Version 3.0", National Remote Sensing Hydrology Program, NWS, Minneapolis, Minnesota, 1988.
- Charbonneau, R., J-P. Fortin & G. Morin, "The CEQUEAU Model: Description and Examples of its Use in Problems Related to Water Resource Management", Hydrologic Sciences Bulletin, 22, 1, 193-202, 1977.
- Clarke, R.T., "Mathematical Models in Hydrology", Irrigation & Drainage Paper 19, FAO, Rome, 1973.
- Fleming, G., "Deterministic Models in Hydrology", Irrigation & Drainage Paper 32, FAO, Rome, 1979.
- Fortin, J.P., J.P. Villeneuve, A. Guilbot & B. Seguin, "Development of a Modular Hydrological Forecasting Model based on Remotely Sensed Data for Interactive Utilization on a Microcomputer", Hydrologic Applications of Space Technology (Proc. Cocoa Beach Workshop, Florida, ed. A.I. Johnson), IAHS Publ. 160, 307-319, 1985.
- Goodison, B.E., I. Rubinstein, F.W. Thirkettle & E.J. Langham, "Determination of Snow Water Equivalent on the Canadian Prairies using Microwave Radiometry", IAHS Publication 155, 163-173, 1986.
- Hatfield, J.L., Perrier, A. & R.D. Jackson, "Estimation of Evapotranspiration at one Time-of-day using Remotely-Sensed Surface Temperatures", Agricultural Water Management, 7, 341-350, 1983.
- Johnstone K.J. & S. Ishida, "An Analogue/Digital Procedure for the Mapping of Snow Cover from Satellite Imagery", Unpublished Manuscript, Atmospheric Environment Service, Environment Canada, 1984.
- Kite, G.W., "Performance of Two Deterministic Models", IAHS Publ. 115, 136-142, 1974.
- Kite, G.W., "Development of a Hydrologic Model for a Canadian Watershed", Can. J. Civil Engineering, Vol. 5, 1978.
- Kite, G.W., "Analysing NOAA Satellite Images from Magnetic Tape", unpublished paper, Canada Climate Centre, NHRC, Saskatoon, 1988.
- Klemes, V., "Conceptualization and Scale in Hydrology", Proc. Int. Symp. On the Understanding of Processes on Basin Scale, Caracas, Venezuela, 1982.
- Morton, F.I., "Operational Estimates of Areal Evapotranspiration and their Significance to the Science and Practice of Hydrology", Journal of Hydrology, Vol. 66, pp.77-100, 1983.
- Moses, J.F. & E.C. Barrett, "Interactive Procedures for Estimating Precipitation from Satellite Imagery", Hydrologic Applications of Space Technology (Proc. Cocoa Beach Workshop, Florida, ed. A.I. Johnson), IAHS Publication No. 160, 25-41, 1986.
- National Research Council, "Global Change in the Geosphere-Biosphere, Initial Priorities for an IGBP", U.S. Committee for an International Geosphere-Biosphere Program, Commission on Physical Sciences, Mathematics and Resources, National Academy Press, Washington D.C., 1986.
- Robinson, A.J. & Associates, "Study of Methodologies of Streamflow Forecasting Incorporating Remotely Sensed Data", Phase I of report for Water Resources Branch, Inland Waters Directorate, Environment Canada, 1986.
- Solomon, S.I., Denouilliez, J.P.; Chart, E.J.; Woolley, J.A. & C. Cadou, "The Use of a Square Grid System for Computer Estimation of Precipitation, Temperature and Runoff", Wat. Res. Res., 4, 5, 919-929, 1968.
- Tschis, A.A. & G.A. Isaac, "On a New Approach for Instantaneous Rain Area Delineation in the Midlatitudes Using GOES Data", J. Climate and Applied Meteorology, Vol. 24, No. 11, 1208-1218, 1985.

TABLE 1
DATA TYPES

| | |
|--------------------|---|
| Basic Station Data | Lat., Long., Grid X,Y Drainage Area |
| Land-use | Coniferous Deciduous Crop Grass No Vegetation Lake Marsh |
| Satellite | Cloud Cover Snow Cover Snow-water Equivalent Surface Temperature Evapotranspiration |
| Physiographic | Mean Elev. Slope Aspect % Basin within Grid Square % Non-contributing Area |
| Climate | Precip. Snowfall Max. Temperature Min. Temperature Radiation |
| Snowcourse | Depth Snow-water Equivalent |
| Hydrometric | Streamflow |

TABLE 2
PRACTICAL CHOICE OF SATELLITE

| | RESOLUTION | FREQUENCY | COST, C\$/IMAGE |
|-----------------|--------------|-----------|-----------------|
| Landsat 5 - TM | 30 m. | 8/16 days | 3840 |
| Landsat 5 - MSS | 80 m. | 8/16 days | 795 |
| SPOT - MSLA | 20 m. | 2 per day | 1855 |
| SPOT - Pan. | 10 m. | 2 per day | 2205 |
| NOAA | 1 km. | 2 per day | 55 |
| GOES | 1 km./ 8 km. | 1/2 hour | 100 |

TABLE 3
SATELLITE DATA AVAILABILITY

| | KOOTENAY | SOURIS |
|--|----------|--------|
| Total no. of images expected (Jan. 1, 1988 - Feb. 28, 1989) | 425 | 425 |
| Total no. of images received | 388 | 397 |
| Total no. of useable images | 378 | 387 |
| No. of images < 100% cloud | 135 | 184 |

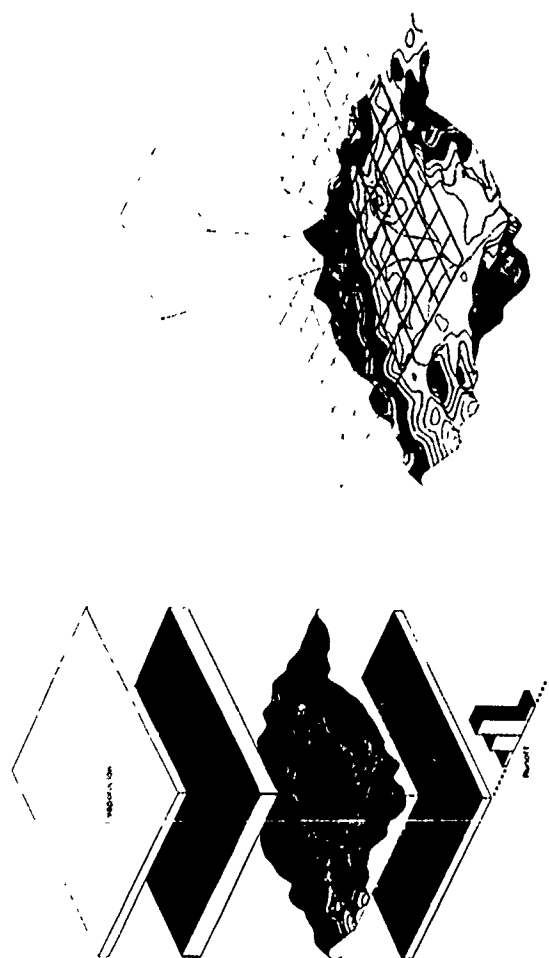


FIGURE 1. LUMPED AND DISTRIBUTED MODELS

DATABASE FOR MODEL STUDIES - TIDEWATER

Station number: Station name:

Year: Month:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----|---|---|---|---|---|---|---|---|---|----|
| Max | | | | | | | | | | |
| Min | | | | | | | | | | |
| Max | | | | | | | | | | |
| Min | | | | | | | | | | |
| Max | | | | | | | | | | |
| Min | | | | | | | | | | |
| Max | | | | | | | | | | |
| Min | | | | | | | | | | |

Next Point: Previous Month: Other Station:

(F1) Help (F2) Normal Exit (F3) ESC Exit (F4) Save

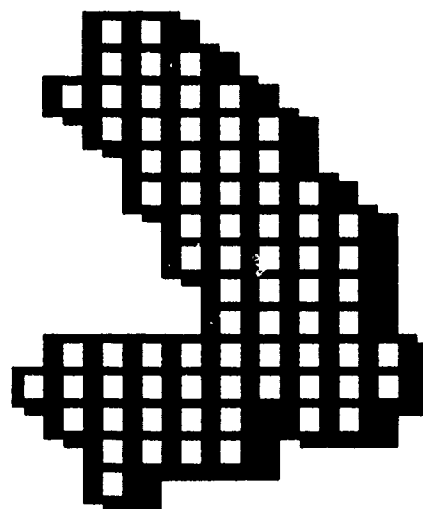


FIGURE 2. DATA ENTRY SCREENS

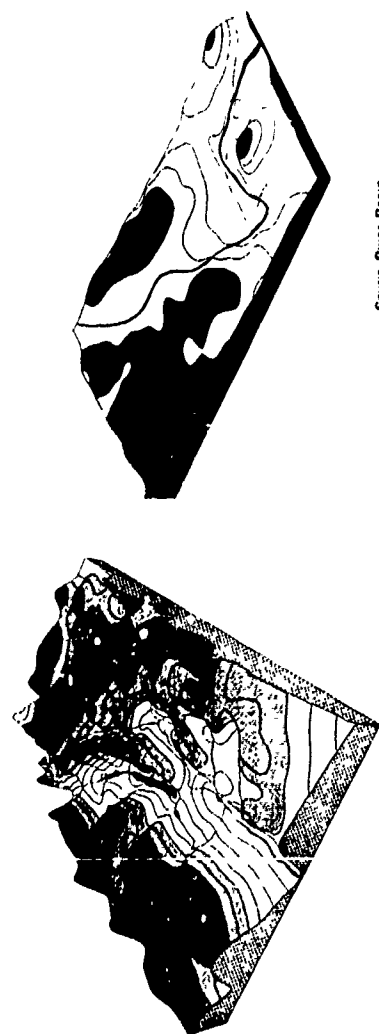


FIGURE 3. 3-D BASIN DIAGRAMS

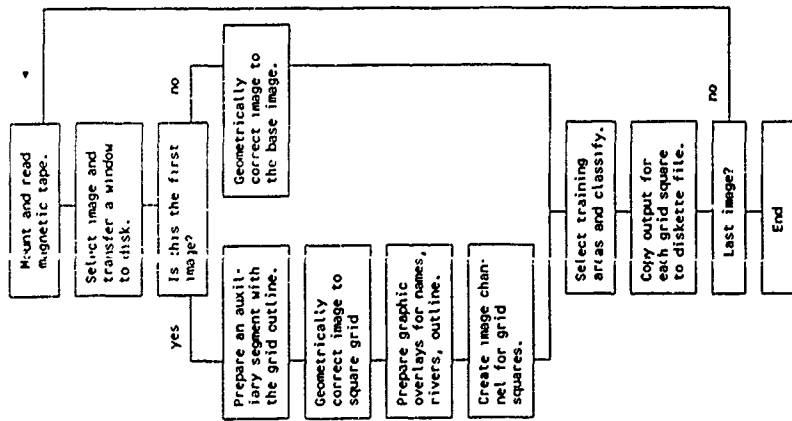


FIGURE 4. FLOWCHART:
IMAGE ANALYSIS FROM MAGNETIC TAPE

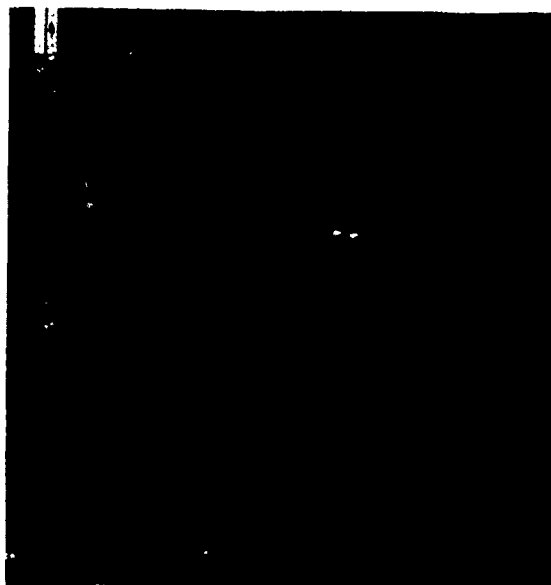


FIGURE 5. NOAA B/W VISUAL & IR IMAGES

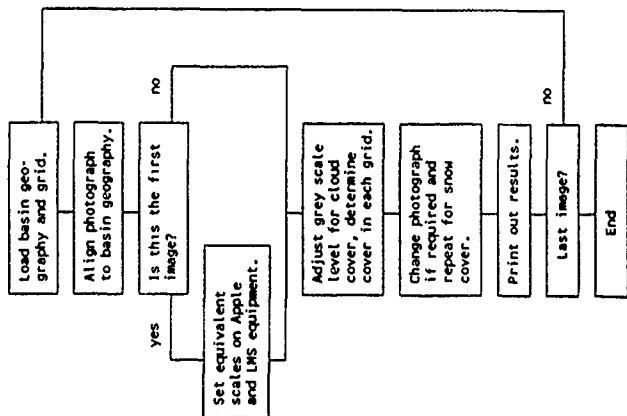


FIGURE 6. FLOWCHART:
IMAGE ANALYSIS FROM B/W PHOTOGRAPH

SNOW COVER MONITORING BY SATELLITES AND REAL TIME RUNOFF FORECASTS

K. Seidel, U. Burkart, R. Baumann

Institut für Kommunikationstechnik ETHZ
Zürich, Schweiz

J. Martinec

Eidgenössisches Institut für Schnee- und Lawinenforschung
Weissfluhjoch-Davos, Schweiz

Abstract

By satellite monitoring of the seasonal snow cover it is possible to evaluate the snow accumulation in mountainous areas. The changing areal extent of snow cover is an important input variable for snowmelt runoff models. Direct measurements of snow reserves are hampered by the difficult access and avalanche hazards in the remote parts of a mountain basin. Therefore remote sensing is being increasingly applied to help to solve these problems. As an example, the paper presents snowmelt runoff simulations for two hydroelectric stations.

Keywords: snow mapping, snow accumulation, runoff forecast, satellites

1. Monitoring of seasonal snow cover by satellites

Earth observation satellites allow to monitor the seasonal changes of the snow coverage in alpine regions. The seasonal accumulation of snow and the gradual decrease of the snow covered area during the snow melt season is a typical feature. Methods are being developed to quantify this process by periodical snow cover mapping.

As an example, Fig. 1 shows the basin of the upper Rhine in the Swiss Alps (*Rhein-Felsberg*; 3250 km², 560-3614 m a.s.l.). The snow cover is shown as it has been evaluated from Landsat-MSS data (Baumgartner, 1987). The areal extent of the snow coverage in different elevation zones can be evaluated for different regions-of-interest or subbasins. If consecutive snow cover maps resulting from different satellite overflights are available, depletion curves can be derived as shown in Fig. 2. In the figure the depletion curves of two smaller subbasins *Sedrun* (108 km², 1840-3210 m a.s.l.) and *Tavanasa* (215 km², 1277-3210 m a.s.l.) are given for comparison concerning the runoff season 1985. The two hydroelectric stations are located within the subbasin *Ilanz* (Fig. 1).

As mentioned, remote sensing capabilities enable us to map the snow cover. Advanced digital image processing techniques are used and with the aid of a digital terrain model (DTM) the snow cover analyses are computed in different elevation zones. By that means the snow cover depletion curves can be plotted from a sequence of such recordings.

2. Evaluation of the areal water equivalent of snow

The depletion curves reflect the seasonal decrease of the snow cover as it is influenced by temperature and precipi-

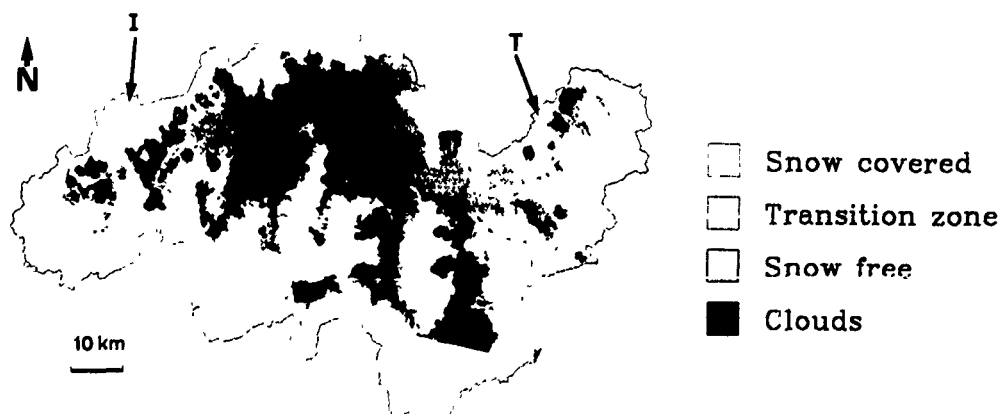


Fig. 1 Basin of the upper Rhine in Switzerland (*Felsberg*) with the snow cover evaluated from the Landsat overflight on 18-MAY-1982. Partial test areas are labeled I = *Ilanz* and T = *Tiefencastel*.

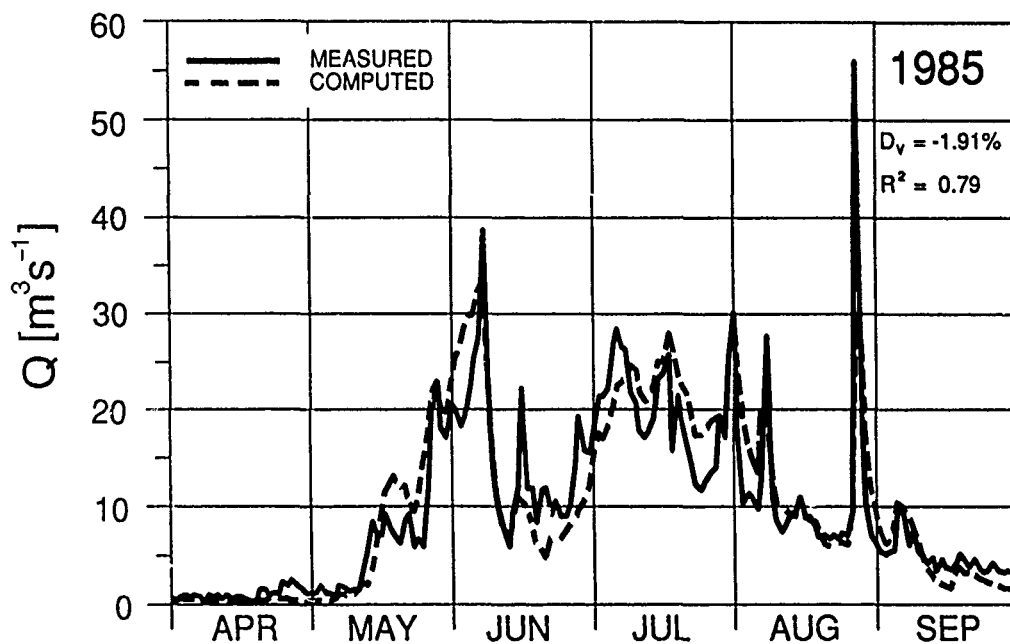


Fig. 4 a) Computed and measured discharge from the catchment area of the hydroelectric station *Sedrun*

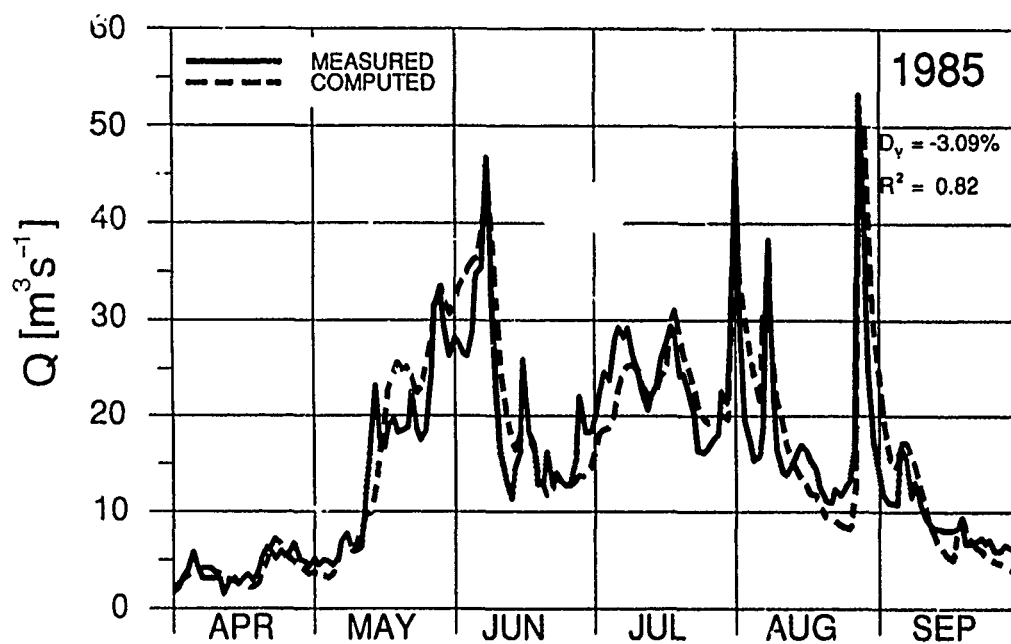
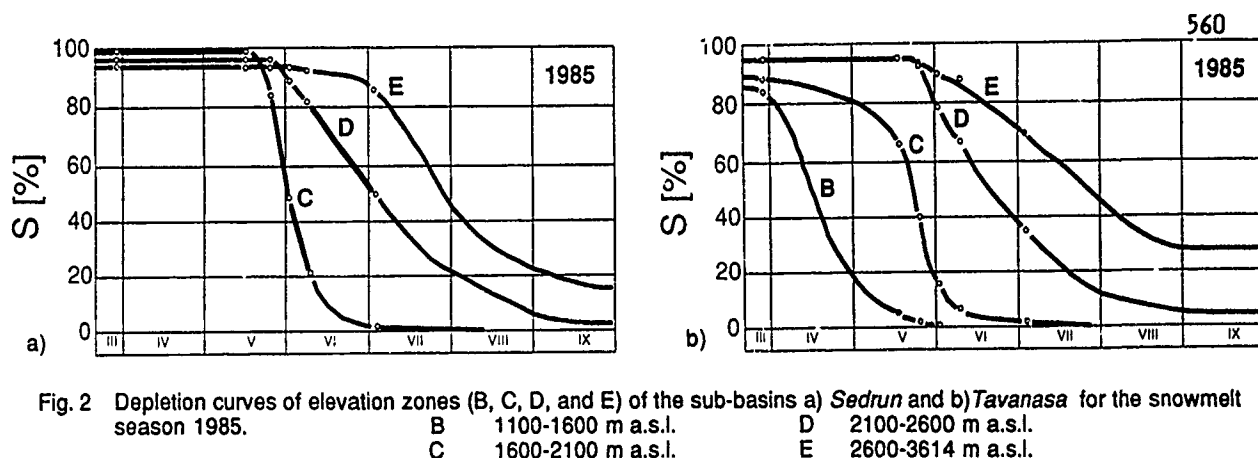


Fig. 4 b) Computed and measured discharge from the catchment area of the hydroelectric station *Iavanasa*.



tation. In order to get the initial accumulation of snow "modified depletion curves" can be derived as explained by Hall and Martinec (1985). These curves relate the areal snow coverage to the cumulative snowmelt depths. In Fig. 3 the modified depletion curves of the subbasins *Sedrun* and *Tavanasa* are shown.

In contrast to a conventional depletion curve of the snow coverage, the modified curve relates the snow coverage with the cumulative snowmelt depth (instead of time). A modified depletion curve thus indicates the initial water volume stored in the snow cover. It is the area between the curve and the x,y-axes. During the snowmelt season, it is also possible to read off the water volume which has flown off and the water volume which is still stored in the snow cover.

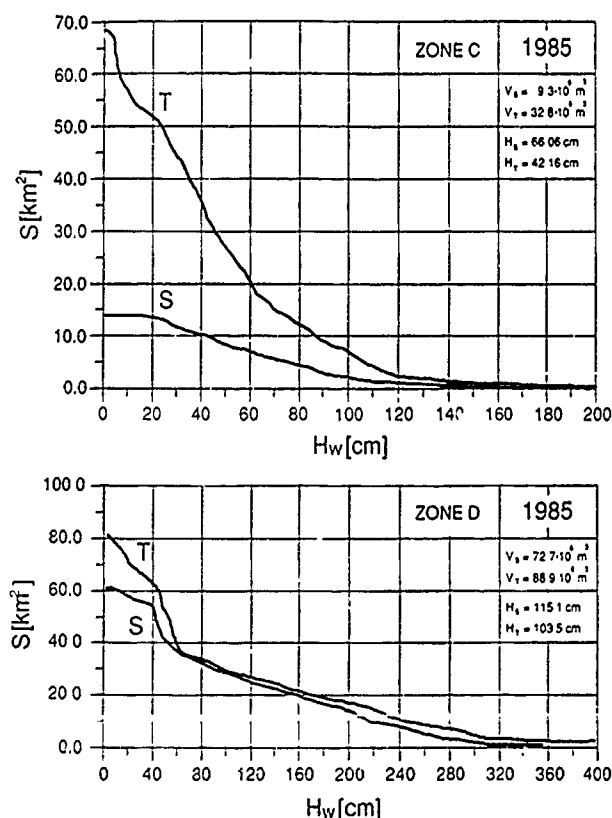


Fig. 3 Modified depletion curves of the snow coverage in the sub-basins *Sedrun* and *Tavanasa* for the runoff season 1985 for comparison in the different elevation zones (C, D, E as in Fig. 2).

This evaluation is applied to the catchment areas of *Sedrun* and *Tavanasa* in Fig. 3. The volume available for *Sedrun* serves for peak electricity production from a reservoir while *Tavanasa* is a run-of-river power plant.

Both informations have been required by an electrical company and found important for an improved production of electricity.

3. Modelling of the snowmelt runoff for hydropower and flood control

The seasonal snow cover is a dominant runoff factor in mountain basins. Recently, the attention has been focused on snowmelt runoff models by a project of the World Meteorological Organization (WMO, 1986). Of the models tested, the SRM model (Martinec et al., 1983) exploits the increasing availability of snow cover mapping from satellites. Besides the air temperature and precipitation, the snow covered area from conventional depletion curves (Fig. 2) is the essential input variable used.

The deterministic snowmelt runoff model (SRM) is designed to simulate or to forecast the daily discharge in mountain basins, resulting mainly from snowmelt but also from precipitation.

Based on the above mentioned satellite snow cover mappings, the daily flows from April to September 1985 have been computed for the previously mentioned hydroelectric stations *Sedrun* and *Tavanasa*. As shown in Fig. 4, the daily runoff values have been simulated for both stations. The model accuracy is characterized by the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^n (Q_i - Q_i')^2}{\sum_{i=1}^n (Q_i - Q)^2}$$

where Q_i is the measured daily discharge
 Q_i' is the computed daily discharge
 Q is the average measured discharge
 n is the number of daily discharge values
 ($n = 183$ in the given case)

and by the volume deviation

$$D_V [\%] = \frac{V - V'}{V} \times 100$$

where V is the measured runoff volume
 V' is the computed runoff volume

No updating or calibration was used, that is to say the model parameters were predetermined and the simulation proceeded on the basis of previously computed discharge. It was thus possible to use snow cover mapping from Landsat data in two relatively small catchment areas.

If such computations can be carried out in real time (using temperature and precipitation forecasts), hydroelectricity production can be increased (Kawata and Kusaka, 1988) and flood control improved. In alpine conditions, floods from snowmelt only are improbable. They occur when the runoff from heavy rainfalls is superimposed on the snowmelt runoff, as illustrated in Fig. 4 by several sharp peaks.

4. Forecasting on the basis of multisensor satellite recordings

In order to be able to use these techniques for (operational) forecasting purposes, one needs "guaranteed information". In the past, usable Landsat-MSS data was not always available for the specified area and the time period under consideration. Based solely on Landsat imagery a runoff forecast cannot be guaranteed.

An obvious consideration is to integrate remote sensing information from other satellite sensors into the snow cover determination scheme as well. It has been discussed by Baumgartner et al. (1988) under what circumstances data from the weather satellite system NOAA/AVHRR can be used for snow cover mapping. It has been reported that due to the spatial resolution of about 1000m x 1000m of NOAA a minimum basin size of about 600-1000 km² is necessary for a digital analysis.

In the meantime data became available from the SPOT-1 satellite and data will become available rather soon from the Japanese MOS-1 system. No major problems have to be expected for the analysis of those data. Due to the fact that they have about the same spatial resolution or even better than Landsat-MSS the resulting snow cover maps will be of adequate quality.

The SPOT system offers in the multispectral mode a ground resolution of 20m x 20m and the ability of its sensors to point up to 27° East and West of the local vertical axis. This latter feature gives interesting possibilities to increase the number

of opportunities to obtain views of a given area. Exactly this property in addition with any further active sensor in orbit prepares the way for an operational procedure for snowmelt runoff forecast.

5. Conclusions

Remote sensing Landsat-MSS data enabled us to map snow coverages even in rather small catchment areas (as requested by a hydroelectric company) in order to serve the requirements of snowmelt runoff modelling.

The paper deals only with Landsat-MSS data, but with the advanced capabilities of the SPOT sensor system and any further earth observation satellite a sufficient frequency of overflights seems to be guaranteed.

Acknowledgement

This paper brings results from a research project at the Institute for Communication Techniques, Swiss Federal Institute of Technology (ETHZ) Zurich, Switzerland. This project is supported by the Swiss National Science Foundation, the Electric Company of Northeast Switzerland (NOK), Telcolumbus Ltd. and by the Swiss Federal Institute for Snow and Avalanche Research at Davos / Weissfluhjoch.

Bibliography

- Baumgartner, M.F., "Schneeschemelz - Abflussimulationen basierend auf Schneeflächenbestimmungen mit digitalen Landsat-MSS- und NOAA/AVHRR-Daten (Snowmelt Runoff Simulations)", Dissertation, University of Zürich, 1987.
- Baumgartner, M.F., K. Seidel and J. Martinec, "Toward Snowmelt Runoff Forecast Based on Multisensor Remote Sensing Information", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE 25, No. 6, p. 746 - 750, 1988.
- Hall, D.K. and J. Martinec, "Remote Sensing of Ice and Snow", Chapman & Hall, London - New York, 1985.
- Kawata, Y. and T. Kusaka, "Snowmelt runoff estimation using snow-cover extent data and its application to optimum control of dam water level", Proceedings of IGARSS '88 Symposium, Edinburgh, Scotland, ESA SP - 284 (Paris), Vol. 1, p. 439 - 440, 1988.
- Martinec, J., A. Rango and G. Major, "The Snowmelt - Runoff Model (SRM) User's Manual", NASA Ref. Publ. 1100, Washington, D.C., 1983.
- WMO, "Intercomparison of models of snowmelt-runoff", Operational Report No. 23, World Meteorological Organization, Geneva, 1986.

RESIDUAL SNOW PATCH MAPPING IN ARCTIC CANADA USING LANDSAT TM IMAGES

¹ JAMES T. GRAY, ² BERNARD LAURIOL, ¹ DENIS BRUNEAU AND ¹ PAUL BRIAND

¹ Département de Géographie, Université de Montréal, Montréal, Québec, H3C 3J7

² Département de Géographie, Université d'Ottawa, Ottawa, Ontario, K1N 6N5

ABSTRACT

Rapid mapping of the areal extent of mid-summer snow patches north of the arctic tree-line, as input to winter snow accumulation models, is readily carried out using LANDSAT TM image data. Discrimination of snow from cloud, bare rock, alluvial gravel beds and beach deposits proved relatively easy, using information combined from TM bands 2, 4 and 5. A linear stretch contrast and training area selection permitted the successful execution of a maximum likelihood classification, followed by measurement of snow-covered areas on all 512 x 512 sub-scenes of the original TM image. Some refinements are anticipated, notably for recognition of snow in shadowed areas, but the technique should prove suitable for large scale snow patch mapping across large sectors of the arctic.

INTRODUCTION

Mapping of snow cover accumulation patterns in Arctic Canada relies heavily on long term snow course data from a very limited number of almost exclusively coastal meteorological stations (Maxwell, 1980). The effects of high topography and continentality may be inadequately represented by such maps, as was shown in a compilation of available data from the Ungava Peninsula (Gray, 1983). Lauriol et al. (1984, 1986) demonstrated by selective measurements of the numbers and areas of residual mid-summer snow patches on air

photographs, that a high correlation existed between snow patch area and the previous winter's snow-fall for the nearest meteorological station. Indeed it appears that for the high arctic, snow patch survival may be more a function of winter snow-fall than of variable snow melt conditions in the spring.

Areal measurements of snow cover, obtained by painstakingly measuring areas of individual snow patches on many hundreds of air photographs, was very time consuming.

Only small sample areas could be assessed, and the periodicity and dates of air photograph coverage were rather variable. The predictable and frequent passage of Landsat 5 over all regions of the arctic, south of approximately latitude 70, and the large areal coverage by single digital images suggested the present evaluation of their potential for residual snow patch mapping.

The main thrust of the present paper is therefore to present a series of digital enhancements of a sample Landsat TM image, obtained from the Canadian Centre of Remote Sensing, as a geometrically corrected, geocoded MOSAICS tape. These enhancements and the ensuing classification permitted discrimination and rapid areal measurement of snow patch extent for a series of 512 x 512 pixel sub-scenes of an image quadrant. The image selected for this analysis (20-17) was obtained for July 13th, 1984 from a topographically varied tundra region in northernmost Ungava (figure 1). A micro-computer based IMAVISION image analysis system in the Department of Geography at the University of Montreal was used for the analysis.

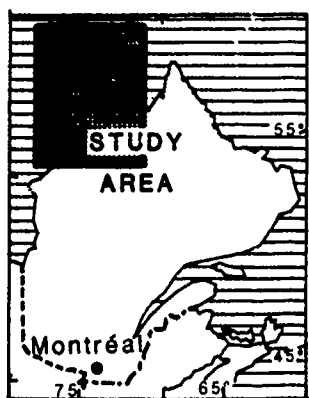


Figure 1. Location of field area

SELECTION OF SPECTRAL BANDS

Various combinations among the six available bands in the visible, near red and middle IR wavelengths were tried, and the best three band combination for discrimination of snow from other light toned zones (cloud cover, alluvial and beach sands and gravels, wave-washed bedrock) was found to be TM 2, 4 and 5. Figure 2 shows the mean spectral response for all six bands, from the snow cover, from a variety of other terrain types and from cloud cover noted on the image. The spectral signature for the snow cover closely resembles the spectral reflectance curves of Choudhury and Chang (1979).

TM 1 (the blue band) shows good contrasts between blue snow cover and the neighbouring vegetation covered tundra, but was not selected on account of atmospheric scattering which reduces surface contrasts in this wavelength. TM 2 (the green band) also shows a strong response for snow covered areas, but is much less affected by the scattering effect. However, confusion is possible between dirty snow patches and non-vegetated beach gravels and sands. Light toned exposed bedrock, such as limestone surfaces could potentially reduce discriminability of snow patches, especially for topographically shaded areas.

Cloud cover also can be confused with snow cover in TM 2. This is a major factor, influencing band selection, some cloud cover being present on practically all arctic imagery. TM bands 3 and 4 in the red and near red IR do not add much to the discrimination already available in TM band 2. They do potentially permit discrimination between wet and dry snow, however. In this study TM band 4 was retained for classification. The mid-IR bands TM 5 and 7 are both characterised by low reflectance values from snow surfaces, and high reflectance values from all other surfaces, except for lakes and rivers. Particularly good discriminability between snow and cloud cover was noted for TM 5. Therefore this band was selected as the third band to be used in image classification.

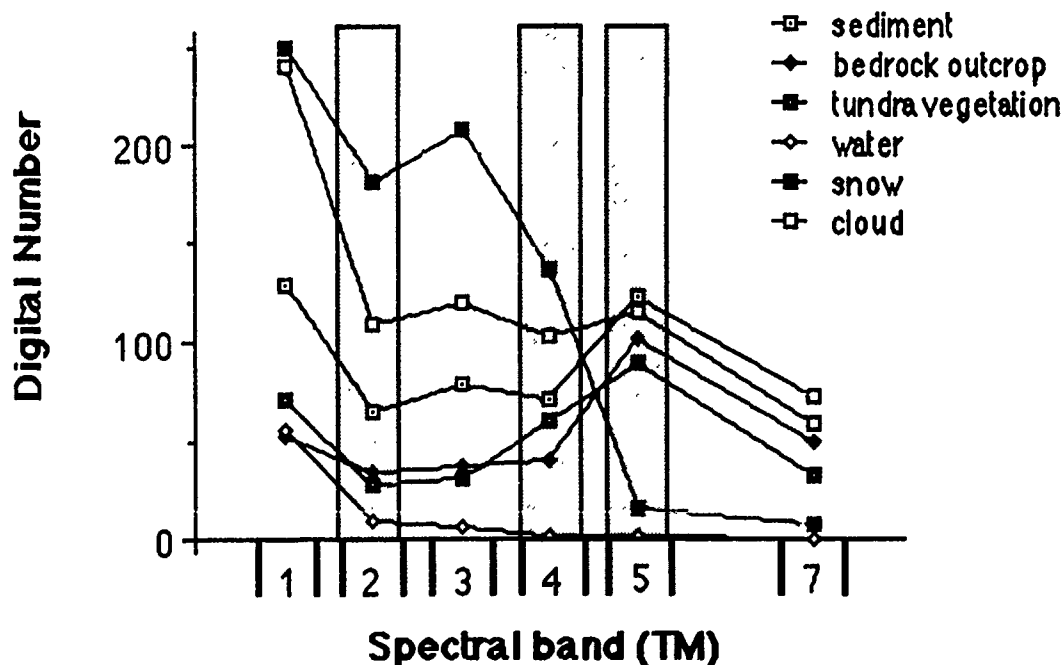
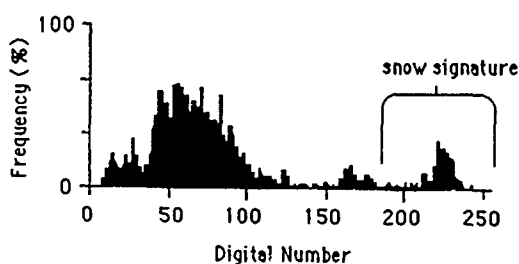


Figure 2. Spectral signatures of snow patches and other terrain types.

IMAGE ENHANCEMENT USING A LINEAR STRETCH CONTRAST

Having selected TM bands 2,4 and 5 for extraction of detailed information on snow cover, the next step is image enhancement in all three bands. The goal of this procedure is to increase efficiency in the selection for classification of training sites. Because the original image has so many relatively low brightness pixels, associated with non-snow covered areas, the snow patches visually appear as rather uniformly bright, almost saturated, zones. It is very difficult to pick out contrasts from the centres to the edges of snow patches. Visual inspection of histograms of spectral response in each band for an entire sub-scene (figure 3) enabled a linear stretch to be carried out on the peak related to snow cover. This peak is stretched out over a range of values from 1 to 255, thus permitting contrasts to be observed between pure snow, dirty snow and mixed pixels at the margins of the snow-banks. These contrasts within the snow covered areas, in fact, lead to the bi-modal distribution of reflectance values shown in figure 4.

a) Histogram TM 2



b) Linear stretch enhancement of snow signature

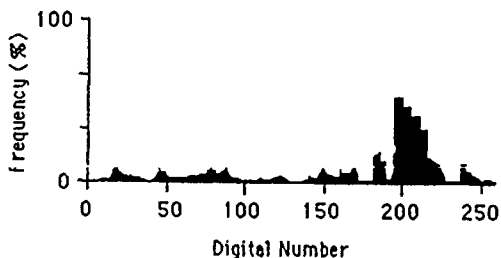


Figure 3. Linear stretch enhancement of snow covered areas.

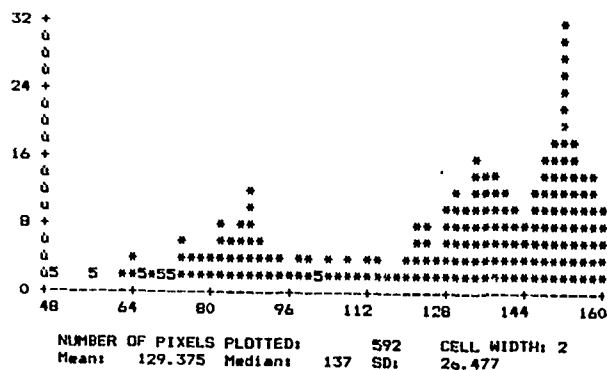


Figure 4. Brightness values for pure and dirty snow.

MAXIMUM LIKELIHOOD CLASSIFICATION

The enhanced image data was then enlarged on the monitor and with the aid of air photographs for comparison purposes, a number of training sites were selected from several snow banks. Each site comprised at least 400 pixels. First order statistical parameters, viz. mean values and variance, were obtained, and a maximum likelihood classification carried out for the entire sub-scene. The total surface area of snow patches was digitally calculated. The classification and areal estimations were then readily applicable to all 512 x 512 pixel sub-scenes of the TM image.

CONCLUSIONS AND FUTURE RESEARCH

Snow patch areas can be rapidly determined for treeless regions of the arctic, using Landsat TM images obtained in mid-summer. Enhancement, classification and areal measurements of such areas were carried out satisfactorily for sample sub-scenes of a TM image in Northern Ungava. The most suitable three band combination for the analysis was TM 2,4 and 5.

Current research on the project is oriented in two main directions. In the first place, some refinement of snow patch recognition techniques is necessary for certain situations. Identification of topographically shaded snow patches in areas of steep relief is difficult because of much reduced brightness values in most TM bands. Also cloud cover precludes the

use of some areas of otherwise good mid-summer TM images. The classification using successive 512 x 512 sub-scenes should permit spatial variability of snow cover to be evaluated for a given image quadrant. This would allow the determination of minimum areas for accurate regional snow cover assessment, thus reducing the problem of modest cloud cover on the image.

In the second place this analysis of mid-summer snow cover will be extended to images, located on a transect from Melville Peninsula to Labrador. The results for several years will then be related to winter snow-fall and winter snow cover, and also to various snow melt parameters measured at nearby climatic station. Ultimately, confidence levels will be established for the relationship between snow accumulation patterns and summer snow patch extent.

ACKNOWLEDGMENTS

This research has been supported by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Ottawa. The northern training grants programme of the Department of Indian Affairs and Northern Development permitted Denis Bruneau to conduct field surveys in 1986 and 1988.

REFERENCES

- CHOUDHURY, B.J. and CHANG, A.T.C. 1979. Two stream theory of reflectance of snow. IEEE Trans. Geosc. Electro, GE-17, 63-68.
- GRAY, J.T. 1983. Extraction and compilation of available temperature and snowfall data in the Ungava Peninsula as input to geothermal modelling of quaternary paleoclimates. Earth Physics Branch, Ottawa. Open Files Report 83-12, 38 p.
- LAURIOL, B., CHAMPOUX, A. and GRAY, J.T., 1984. Répartition estivale des surfaces enneigées en Ungava, Nouveau Québec. Géographie Physique et Quaternaire 36(1): 37-48.
- LAURIOL, B. BEAUDET, H. CARRIER, Y. and BINDA, G. 1986. Spatial and temporal variations of the residual snow cover in the Canadian Arctic in July, 1948 to 1983. Arctic 39 (3).
- MAXWELL, J.B., 1980. The climate of the Canadian Arctic Islands and adjacent waters: Vol.1: Climatological Studies No. 30, 531 pp.

THE TECHNOLOGICAL CHALLENGE OF FUTURE SPACE RADAR SYSTEMS

by

Hans Martin Braun, Gerhard Hans Rausch, and Helmut Kappel
Dornier GmbH, Postfach 1420, 7990 Friedrichshafen

ABSTRACT

Today, over 10 years after the operation of the famous SEASAT sensors synthetic aperture radar (SAR) and wind scatterometer (SCATT), the time has come for the next generation of such radar systems. The last two decades were dominated by extensive technological developments in order to build spaceborne radars. However, demanding user requirements on operational space radars to be flown on the forthcoming polar platforms are calling for a further increase in technological capabilities. This paper discusses technological requirements for future civil space radar programmes, shows present developments within the key technology areas and ends up with recommendations on the most important areas of investigations in the near future. This review is based on the European and German programmes of Earth Observation and Basic Technology Research.

Keywords: SAR Technology, microwave scatterometer technology, active array, CFRP antenna, power amplifier, digital pulse generation, SAR processing.

INTRODUCTION

Radar remote sensing from space was a key element in international space programmes for more than a decade now. It even becomes more important today when problems within our natural environment are calling for global observations of environmental effects on a routine basis. Spaceborne radars, such as Synthetic Aperture Radars and Microwave Scatterometers, are viable tools for operational observation-systems because they can operate night and day and usually independent from weather conditions.

DORNIER was involved in the development of spaceborne radars from the beginning of these programmes on behalf of the European Space Agency ESA and the German Ministry for Research and Technology BMFT. Today, DORNIER is leading the development of the ESA Earth Observation Satellite ERS-1 and the X-Band Synthetic Aperture Radar X-SAR for Shuttle, the latter being a joint programme with NASA's Shuttle Imaging Radar SIC-C. Furthermore, DORNIER will be responsible for the procurement of the instruments to be flown on the first mission of the European Polar Platform PPF and for the development of the advanced X-SAR sensor for the EOS.

Within the framework of the polar platforms the status of radar remote sensing from space will change from experimental to operational. This implies that versatile radar instruments have to be reliable over a long lifetime with constant measurement performance. They shall represent modern remote sensing techniques and they shall be flexible enough to allow the adaption of their operations to various user needs by changing (reprogramming) modes and parameters in orbit. This flexibility requires the implementation of new techniques and technologies in the areas of radar antennas, power amplifiers, radar pulse generation, and onboard data processing. Following this line DORNIER focused technological developments to

- light-weight array antennas
- pulse radar amplifiers and highly efficient power conditioning systems
- digital pulse generators
- onboard real-time SAR processors

All of these areas need innovative designs, research in new materials, and investigations in new techniques and processes in order to provide light-weight systems with low power consumption and high reliability fulfilling the demanding space qualification requirements.

ADVANCED SYSTEMS

The development of spaceborne radar systems is in progress along the following lines.

- Preoperational/operational systems
- Experimental systems

On the one hand there are systems to be flown on preoperational/operational Earth observation satellites. Although they aim at ambitious objectives from a technological point of view, one can consider them to be relatively conventional from a conceptual and scientific point of view. Radars of this type will be flown on the ESA remote sensing satellite ERS-1, the Canadian RADARSAT, the Japanese satellite JERS-1 etc..

Advantages of such systems are their long term operational capabilities and their world-wide availability for practical use in the near future. Apart from the 100 days mission of SEASAT in 1978 there has not been any such opportunity yet.

On the other hand, experimental systems have been (and will be) developed and operated, which serve to promote the basic knowledge of microwave remote sensing, sensor capabilities and system aspects. The most prominent projects of this type in the past were the Shuttle Imaging Radar systems SIR-A and SIR-B and in the near future it will be the first multifrequency SAR system in space, the SIR-C/X-SAR mission being a joint effort of the Jet Propulsion Laboratory and Dornier. Sensors of this type are forerunners of the next generation of space radars. The user requirements on such advanced imaging radars (SARs) are developing towards multifrequency, dual-polarization, and flexibility (high resolution/medium swath vs. medium resolution/wide swath).

Future requirements on advanced microwave scatterometers are mainly concerning measurement sensitivity and stability at reduced levels of primary power.

The next application opportunities for these advanced sensors are the polar platforms to be launched at the end of this century. The technological competition towards them has already been started.

ANTENNAS

Spaceborne radar sensors like Synthetic Aperture Radars and Microwave Scatterometers typically need very large array antennas. Two different technologies are presently in use:

- Microstrip patch arrays
- Slotted waveguide arrays

The 9.40m x 2.10m SAR antenna of SEASAT was a microstrip patch array for L-Band. Due to the relatively high feeding losses of this antenna type, which even increase with frequency, ESA decided in 1980 to realize the ERS-1 C-band SAR antenna (Fig. 1) as a slotted waveguide array. Because of the high mass and the unsatisfying thermal stability of such an antenna made from aluminium, copper or brass, the novel technology of metallized carbon fibre reinforced plastics (CFRP) was used. Today, this new technology is established within the ERS-1 programme; the electrical performance is as high as for metal antennas, the weight is about 1/3 and the thermal stability and the stiffness is very high. This success led to the decision to apply this new technology within the German

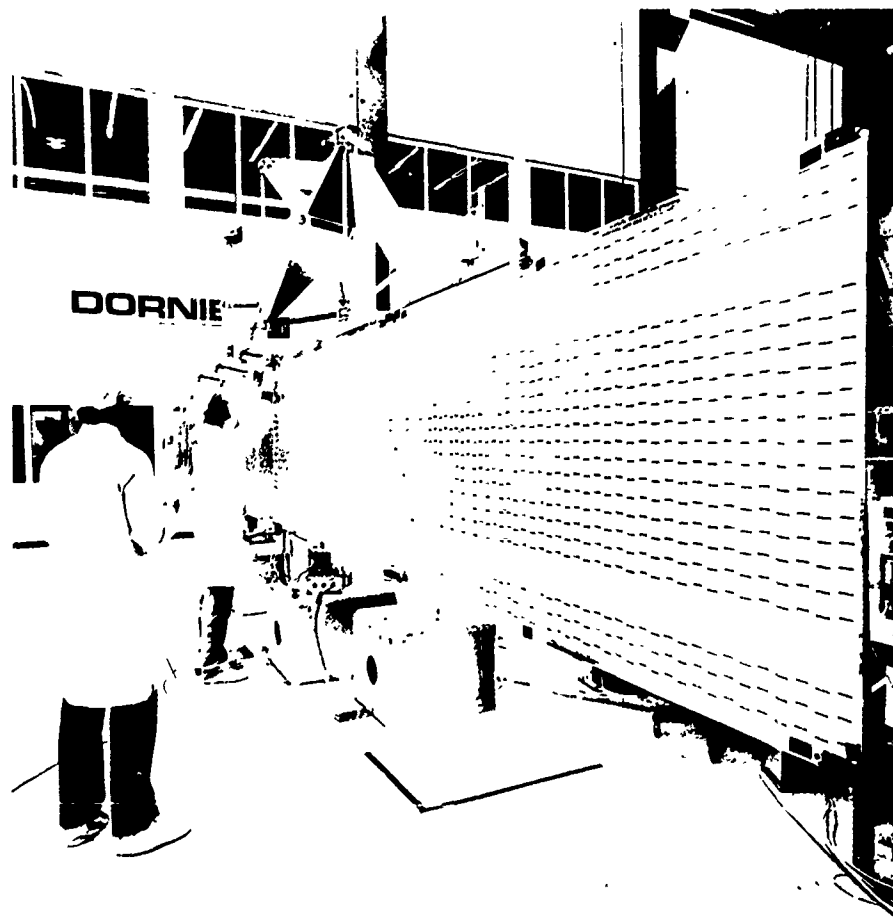


Fig. 1: The ERS-1 SAR antenna development model

Shuttle X-band SAR (X-SAR). Fig. 2 shows a breadboard of an X-band SAR antenna panel. The X-SAR is the German contribution to NASA's Spaceborne Radar Lab (SRL) operating in 3 frequencies (L-, C- and X-band) with two polarizations for L- and C-band. The L- and C-band channels of the SRL are using active phased array antennas with distributed solid state amplifiers. The radiating surface is a microstrip patch structure. This antenna represents the next step into the technological future of spaceborne radar antennas being the distributed amplifiers phased array technology. Feeding of this antenna is performed on low signal level. Many solid state amplifiers at the rear sides of the electrical subpanels provide the required peak power, and phase shifters within these amplifiers allow the steering of the antenna beam. Switches select the required polarization. This active array of SIR-C is still a high-mass-solution and the structural and thermal problems are evident. This is acceptable for a Shuttle experiment, however, additional innovations are needed in order to lower the mass and to solve thermal problems before such an antenna can be applied to an operational SAR like the Facility-SAR on EOS.

Microwave Scatterometers require a high accuracy and stability of the electrical parameters, because of the very high sensitivity of the wind data extraction algorithms in terms of radiometric performance. CFRP slotted waveguide antennas meet these requirements, whereas in-flight stability and calibration of active phased arrays are still facing problems and hence there are some doubts whether active arrays are presently able to meet these requirements. Therefore it makes sense to stay with CFRP slotted waveguide arrays for the next generation of microwave scatterometers.

In order to solve some of the essential problems associated with the active array technology ESA stated a development programme in 1987. These first investigations covered breadboarding of the basic functions in conventional technology in order to prove the functionality of the intended operational modes (wide-beam transmitting, dual-beam receiving, receive-scanning). The transfer to space-qualified units will begin this year. In view of ESA's advanced C-band SAR this development deals with the C-band technology.

In Germany an active array is needed for the X-SAR follow-on project, the XEOS, a possible German contribution to NASA's FACILITY SAR. This is an X-band radar and hence an X-band active array technology programme will likely be started within this year in Germany.

POWER AMPLIFIER

The discussions on the optimum antenna concept are directly related to the investigations into the technology of high microwave power generators for pulsed radars, which allow to maintain the signal coherency. Up to now, mainly travelling wave tube amplifiers (TWTAs) have been used. In Germany, on behalf of the Federal Ministry for Research and Technology, Dornier and AEG developed space qualified high power TWTAs for C- and X-band in order to serve the ERS-1 Active Microwave Instrumentation (AMI), operating in SAR and SCATT mode, and the X-SAR. An application of such an TWA to the Canadian RADARSAT is also under consideration. The technological challenge within these developments was to achieve a high electrical efficiency, low mass, and, at the same time, to solve the essential problems concerning thermal dissipation and high voltage/power discharge under space environmental conditions.

However, as briefly indicated before, the next generation of spaceborne radar antennas, the distributed amplifiers phased arrays, are initiating the new technology of high power solid state pulse amplifiers. These amplifiers (a breadboard is shown in fig. 3) are designed as moduls with an output power in the order of 10 - 20 Watts. Located on the rear side of large planar arrays they typically feed small antenna subarrays or even single radiating elements. Problems to be solved in this development are mainly concerned with electrical efficiency, thermal dissipation, radiation hardness, lifetime, phase and amplitude stability, and mass.

PULSE GENERATION

High bandwidth phase coded radar pulses today are mainly generated by surface acoustic wave devices. Typical values of rf-bandwidth are 10 - 30 MHz. In some applications (e.g. radar altimeters) even 400 MHz are obtained by special signal processing techniques. In case of modern radar systems adaptive pulse phase coding is required. This means, that it must be possible to change the bandwidth, the compression ratio, the phase slope, and the amplitude slope of the pulse by commands. This allows to adapt the spatial resolution of the radar to various users' needs. This requires either series of surface acoustic wave devices or a programmable digital chirp generator (DCG). Intensive developments of DCGs are presently ongoing in Europe and in the US. Bandwidths of more than 100 MHz have already been obtained in experimental systems.

ONBOARD DATA PROCESSING

High-speed real-time onboard radar data processing is required in operational systems in order to lower the data transmission rate. Radar altimeters need processing of signals of 300 - 400 MHz bandwidth and Synthetic Aperture Radars provide data streams of 100 Mbps and more which should be processed in real-time. The operations baseline of today's systems is the processing of radar altimeter data onboard and SAR data on ground. In future the planned operation of several SAR modes in parallel (e.g. 4 polarizations, 3 frequencies) drastically increases the overall data rate (> 500 Mbps). This is beyond the capacity of existing downlink channels and therefore either a limited number of channels have to be selected for transmission or the SAR data must be processed onboard into multilook images providing a significant reduction in data transmission rate. Dornier has developed a real-time SAR pipeline processor, presently being used as quick-look processor on ground. Its design was selected towards a possible application onboard future advanced space SARs considering low power consumption, low mass, and low volume.

OUTLOOK

Operational radar remote sensing from space will be a key element in Earth monitoring at the turn of the century. Even the remotest areas on Earth, most inaccessible to man, will be surveyed on a routine basis, independent of daylight or cloud cover. Users from different scientific disciplines, such as geology, oceanography, hydrology, cartography, vegetation science and many others, require demanding measurement accuracy leading to the application of sophisticated technologies. Industry is going to take this challenge.

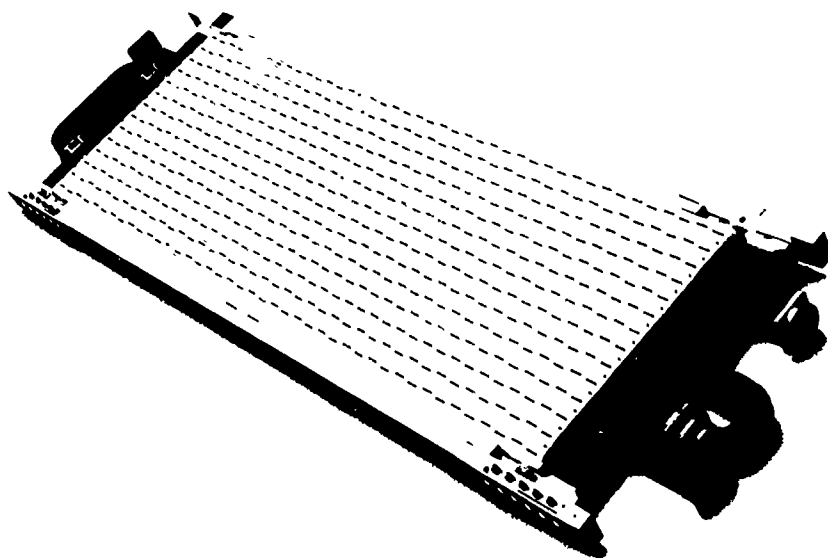


Fig. 2: The X-SAR antenna panel breadboard

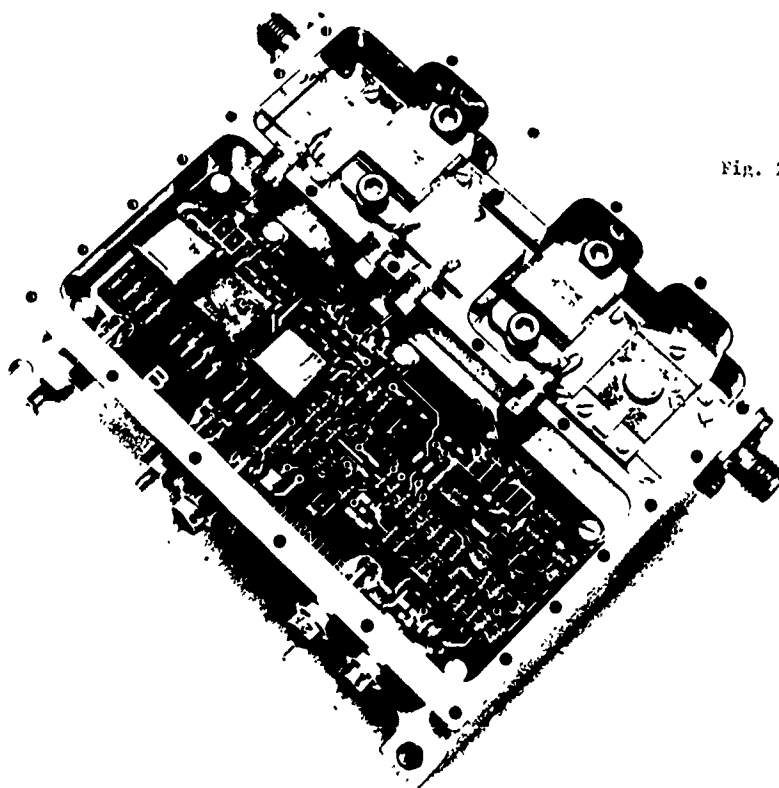


Fig. 3: The C-band solid state amplifier breadboard

Multimission Capability of the European Polar Platform

R. Benz, G. Gebauer, F. Tanner and E. Zeis¹

Dornier GmbH, Postfach 1390, D-7990 Friedrichshafen, FRG

¹) ESTEC, P.O. Box 299, 2200 AG Noordwijk, The Netherlands

Abstract

Under contract to the Earth Observation Preparatory Programme of the European Space Agency a multimission and programme aspect study has been performed for the two European Polar Platform concepts currently under study in the Columbus programme.

For these two different platform concepts various potential payload complements comprising in total more than 25 different instruments have been accommodated in order to prove the multimission capability of both platforms, making use of commonality either to the other Columbus elements or to the SPOT programme. The payload complements are built up by an operational instrument package for meteorological applications, by so called core facility instruments for Earth observation, by Announcement of Opportunity (A.O.) and space science instruments. The payload complements have been seen as typical examples for ocean/meteorological/climate missions and for land missions to cover at least four missions of the Polar Platform. Both platform concepts could demonstrate in principle their flexibility to accommodate the required payload complements. Different payload module capabilities as well as the different flight orientation of both platforms and dissimilar payload accommodation principles resulted in two distinct payload module solutions.

As a further major study result, the importance of physical and functional platform modularity could be substantiated with special attention to the AIV programme under multimission aspects.

1. INTRODUCTION

In parallel to the C-Zero system studies for the European Columbus Polar Platform (Jenkin, R. et al., 1989; Cornet, J. et al., 1989), a multimission and programme aspects study has been conducted by the Earth Observation Preparatory Programme to prove the multimission capability of the proposed platform concepts and to support the consolidation of mission and system requirements (Benz, R. et al., April 1989). This reflects that the Columbus programme aims at developing a single Polar Platform, while the Earth Observation Programme considers a series of several missions based on recurrent models of this platform. The presented study has recently run into a mission phase A system study which includes the design of the core facility instruments in order to achieve finally

a consistent and reliable payload complement definition.

2. MISSION ASPECTS

The European Polar Platform missions shall continuously provide a wide multidisciplinary community with synoptic and coherent data on global and local scale, ranging from fully operational applications for meteorological services with rather short timescale response requirements to purely scientific objectives. Consequently the payload complements consist of operational instruments, developed and funded by operational agencies like NOAA and EUMETSAT, core instruments, developed and funded by ESA and, last but not least, of space science and A.O. instruments funded either by ESA or national agencies. The total number of instruments which can be flown in each mission, is about 12 to 15, dependent on the selected platform concept.

The payload complements comprise optical, microwave, particle and field instruments as well as a direct broadcast package for the operational applications. Consequently, field of view, sun illumination, pointing accuracy and stability, EMC, thermal control, commanding and control, and other requirements had to be studied and optimized carefully, in order to achieve an acceptable compromise for all instruments, without driving the platform capabilities unnecessarily (see Benz, R. et al., Jan. 1989).

In particular, orbit requirements have been deduced from various, partly diverging instrument requirements, resulting in a sunsynchronous, near polar orbit of an altitude between 700 and 850 km, with global coverage for operational instruments for latitudes above about 40°, suitable ground track pattern, and a nodal crossing time with acceptable solar illumination conditions for the optical instruments. It is evident that the currently chosen nominal orbit parameters, as presented in Tab. 1, could only be a compromise and should be further optimized.

E.g., the chosen altitude was mainly requested by the operational instruments, which would fly even higher. Other orbit parameters, e.g. lower orbit altitude of about 762 km or later nodal crossing time will become possible for the land missions in the parallel mission scenario because of the absence of the operational instrument package.

| | |
|----------------------------------|-----------|
| Sunsynchronous, near polar orbit | |
| Altitude | 824 km |
| Inclination | 98.7° |
| Argument of perigee | 50° |
| Local time (descending node) | 9.5 h |
| Duration of repeat cycle | 14 days |
| Nodal period | 101.3 min |

Tab. 1: Current Nominal Orbit Parameters

Two different mission scenarios have been defined for the two platform concepts: The interleaved scenario has been applied to the larger platform option A, while the platform option B is foreseen for the parallel mission scenario. This is not a necessary, but a realistic consequence of the smaller payload mass capacity of option B. The launch sequence of both mission scenarios is given by Fig. 1.

Under the given programmatic constraints, both mission scenarios offer a very similar applicational and scientific benefit. However, operational continuity of the meteorological instrument package seems more critical in the parallel scenario in case of a launch or early operation failure.

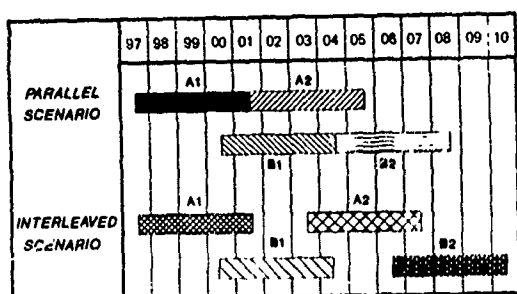


Fig. 1: Alternative Mission Scenarios

3. PAYLOAD COMPLEMENT BUDGETS

For each platform concept a set of 5 payload complements has been defined (see Tab. 2). The payload complements for the first missions are typical for ocean/meteo/climate purposes, while the complements for the second missions are typical for land applications. The operational continuity for the meteorological applications must be guaranteed with one mission type in case of the parallel scenario. Consequently the operational instrument package is not required in the option B land missions. Besides that, for the larger option A platform the rule 2 of 3 has been applied for the core instruments, which means that 2 of the 3 instruments SAR, ATLID and MIMR must be accommodated in each ocean/meteo/climate mission. For option B the rule 1 of 3 has been applied accordingly.

Tab. 3 demonstrates that both platform concepts provide sufficient margin to carry all the payload complements from mass point of view, including the extra mass for interface hardware, as support panels or even complex deployment mechanisms.

For payload power consumption and data rates the situation is more complex, because realistic operational

| Instruments | OPTION A | | | | | OPTION B | | | | |
|-----------------|----------|----|----|----|----|----------|----|----|----|----|
| | 1a | 1b | 1c | 2a | 2b | 1a | 1b | 1c | 2a | 2b |
| AMRIR | x | x | x | x | x | x | x | x | | |
| AMSU A+B | x | x | x | x | x | x | x | x | | |
| ERBI-NS | x | x | x | x | x | x | x | x | | |
| ARGOS | x | x | x | x | x | x | x | x | | |
| SARSAT | x | x | x | x | x | x | x | x | | |
| DB | x | x | x | x | x | x | x | x | | |
| ALT | x | x | x | x | | x | x | x | | x |
| SCATT L+R | x | x | x | | | x | x | x | | |
| MERIS | x | x | x | | | | | | x | x |
| MIMR | x | | x | | | x | | | | |
| SAR | | x | x | | | | | x | | |
| ATLID | x | x | | | | | x | | | |
| HRIS | | | | x | | | | | x | |
| HRTIR | | | | x | x | | | | x | |
| SARA | | | | x | x | | | | x | x |
| LISA | | | | | x | | | | | x |
| VHROI | | | | | x | | | | | x |
| SEM | x | x | x | x | x | x | x | x | | |
| AURIO | x | x | x | x | x | x | x | x | x | x |
| GEM* | x | x | x | x | x | x | x | | x | x |
| PAF | x | x | x | x | x | x | x | x | x | x |
| FPI | x | x | x | x | x | x | x | x | x | x |
| A.O.Instruments | x | x | x | x | x | x | x | x | x | x |
| PDHT | x | x | x | x | x | x | x | x | x | x |

- 1 : ocean/meteo/climate missions
 2 : land missions
 * : may not be accommodated in option B
 a,b,c: mission options

Tab. 2: Payload Complement Options

scenarios have to be taken into consideration. With dedicated operational requirements for each instrument modelled according to their zones of interest (see Tab. 4) and two different realtime transmission scenarios, power timelines have been established. The first transmission option assumed that instruments with data rates >3 Mbps are operated only during ESA station contact, while the second scenario assumed realtime operation during contact with the two DRS satellites, resulting in significantly different instrument duty cycles (see Tab. 5). The two realtime transmission scenarios must be seen as extreme examples for duty cycles of high data rate instruments, allowing for further optimization to achieve the final payload operations scenario. E.g. operating a SAR instrument about 38 % in image mode or science instruments more than 40 % of the orbit seem to be unnecessarily high and could be optimized further. All instruments not mentioned in Tab. 5 operate continuously. Their data are recorded on board and dumped to Earth during ground contact with Kiruna or an other ESA station.

The resulting payload power consumption, as summarized in Tab. 6, is relatively high for both platform concepts and both transmission scenarios compared to the given requirement. While the marginal power consumption, in particular during eclipse phases, can likely be solved by tailored payload operations without limiting the user needs in fact, the high peak power can hardly be solved by modified operations. Here the platform capabilities should be improved appropriately.

| Payload Complement | Option A | | | | | Option B | | | | |
|--------------------------------|----------|------|------|------|------|----------|------|------|------|------|
| | 1a | 1b | 1c | 2a | 2b | 1a | 1b | 1c | 2a | 2b |
| Net Instrument Mass (kg) | 1623 | 1673 | 1623 | 2035 | 2245 | 1265 | 1315 | 1515 | 1508 | 1768 |
| Mechanical Interface Mass (kg) | 247 | 253 | 300 | 224 | 244 | 185 | 166 | 168 | 150 | 163 |
| Gross Payload Mass (kg) | 1870 | 2126 | 2123 | 2259 | 2489 | 1450 | 1481 | 1683 | 1658 | 1931 |
| Margin (%) | 33.2 | 24.1 | 24.1 | 19.3 | 11.1 | 27.5 | 26.0 | 15.9 | 17.1 | 3.5 |
| Utility Module Capability (kg) | 600 | | | | | 2000 | | | | |

Tab. 3: Payload Complements Mass Budgets

| Zones of Interest | Instrument | Remarks |
|------------------------|--|--|
| Continuously operating | NOAA package, ALT, ATLID, FPI, LISA2, MIMR, SEM | |
| Ocean | SAR Scatt | wave mode 1 of 6 antennas, min. duration >180 min |
| Coast | HRIS ⁽¹⁾ SAR ⁽¹⁾ | only during day image mode |
| Land | HRIS ⁽¹⁾ HRTIR ⁽¹⁾ SAR ⁽¹⁾ SARA ⁽¹⁾ VHROI ⁽¹⁾ | only during day image mode only during day image mode only during day image mode |
| Ice | HRTIR ⁽¹⁾ SAR ⁽¹⁾ | only during day image mode |
| Altitude Bands | AURIO ⁽¹⁾ GEM ⁽¹⁾ PAF | high mode, else low mode |
| Day | MERIS ⁽¹⁾ | |

(⁽¹⁾ only operated under realtime transmission)

Tab. 4: Instrument Zones of Interest

| | ESA Stations | | | DRS Coverage | | |
|-----------|--------------|-------|-------|--------------|-------|-------|
| | Day | Night | Total | Day | Night | Total |
| MERIS | 2.5 | - | 2.5 | 25.2 | - | 25.2 |
| SAR IMAGE | 5.7 | 1.3 | 7.0 | 30.7 | 8.2 | 38.9 |
| SCIENCE | 6.0 | 0.1 | 6.1 | 37.1 | 5.1 | 42.2 |
| HRIS | 3.3 | - | 3.3 | 14.9 | - | 14.9 |
| HRTIR | 3.4 | - | 3.4 | 19.8 | - | 19.8 |
| SARA | 2.7 | 0.9 | 3.6 | 16.2 | 6.3 | 22.5 |
| VHROI | 3.2 | - | 3.2 | 14.8 | - | 14.8 |

Tab. 5: Duty Cycles for Intermittently Operating Instruments (in percent)

Significant differences between the ocean/meteo/climate missions and the land missions became apparent for the data rates. While the first mission type requires about 2 x 100 Mbps, the land missions demand for more than 4 x 100 Mbps. This indicates again the

| | ESA Stations Coverage | | | DRS Coverage | | |
|-----|-----------------------|-----------|-----------------|--------------|-----------|-----------------|
| | Day (W) | Night (W) | Orbit Aver. (W) | Day (W) | Night (W) | Orbit Aver. (W) |
| A1a | 1617 | 2586 | 2607 | 2792 | 2644 | 2748 |
| A1b | 2766 | 2717 | 2751 | 3278 | 2994 | 3193 |
| A1c | 2196 | 2147 | 2181 | 2708 | 2424 | 2623 |
| A2a | 2124 | 2037 | 2098 | 2644 | 2369 | 2562 |
| A2b | 2270 | 2191 | 2246 | 2760 | 2523 | 2690 |
| B1a | 1475 | 1467 | 1473 | 1519 | 1484 | 1509 |
| B1b | 2045 | 2036 | 2042 | 2089 | 2054 | 2079 |
| B1c | 1624 | 1598 | 1616 | 2006 | 1833 | 1954 |
| B2a | 1353 | 1288 | 1334 | 1892 | 1620 | 1811 |
| B2b | 1641 | 1592 | 1626 | 2117 | 1924 | 2059 |

(Actual System Requirement: Day 2700 W, Night 2300 W)

Tab. 6: Payload Complements Net Power Consumption (averaged over 15 orbits)

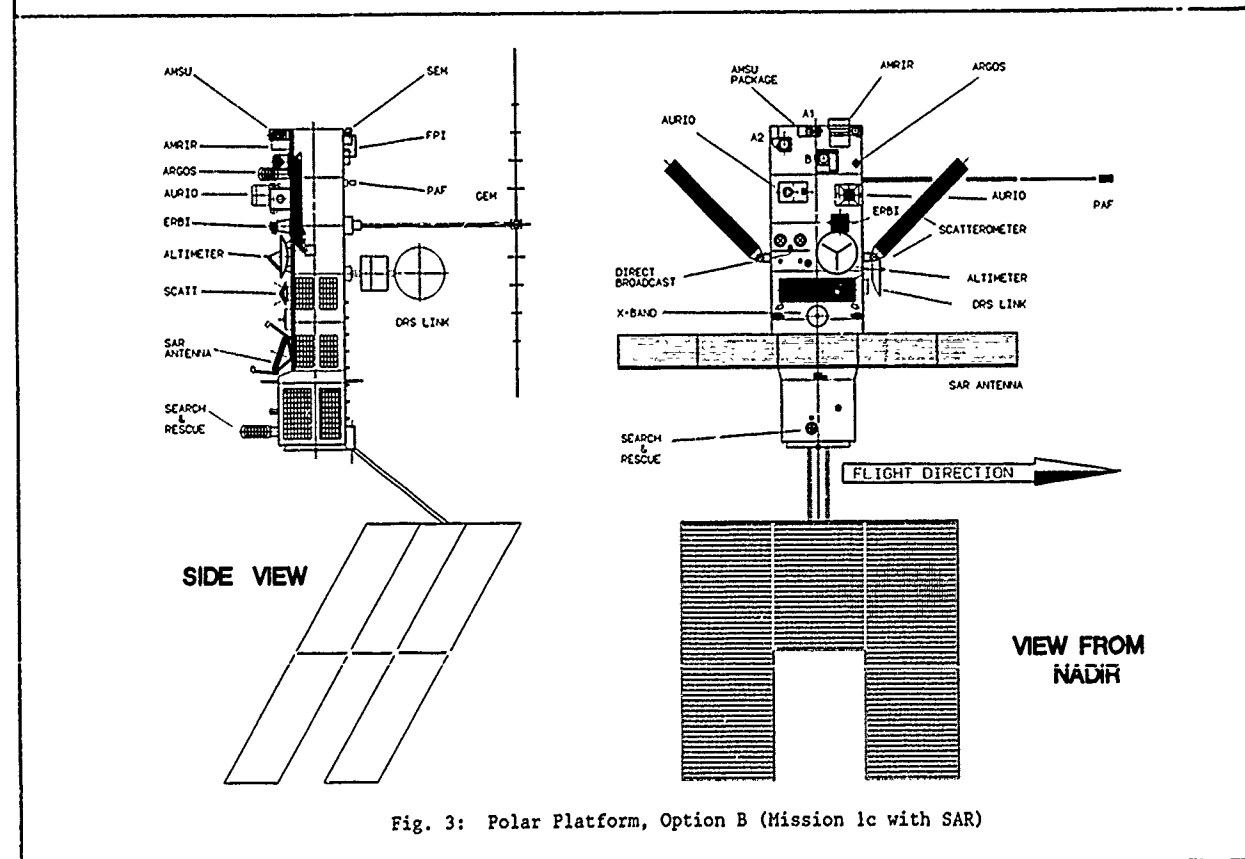
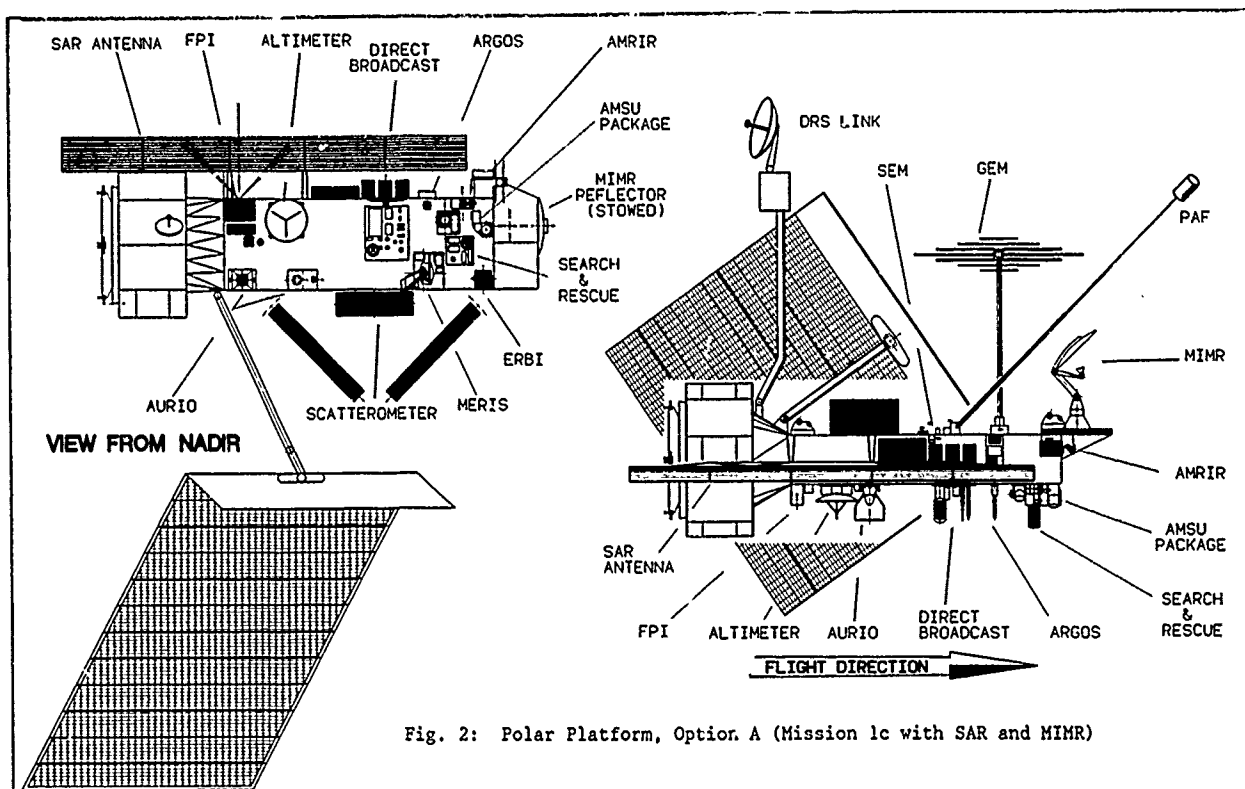
need to optimize the payload composition and the operation scenario further to achieve a suitable balance between user needs and realistic platform capabilities.

4. PLATFORM CONCEPTS AND PAYLOAD ACCOMMODATION

A brief comparison of both platform concepts under payload aspects is given in Tab. 7. Distinct variations in payload accommodation are caused by the different flight orientation of the platforms. While option A is oriented in the flight direction, option B flies transversal, which eases the accommodation of the solar array and of critical instruments, in particular SAR and MIMR. Consequently the effort for instrument deployment is larger for option A. Another important difference is the accommodation of the payload electronics. While in option A all electronic units are attached to the outside of the payload carrier structure, with substantial advantages for the instrument integration, option B houses the instrument electronics together with the mission peculiar payload support functions in the payload equipment bay. In summary a proper accommodation of all required payload complements could be demonstrated. Some critical thermal control and EMC aspects seem to be solvable during the subsequent detailed design phases. Fig. 2 and 3 illustrate the configurational characteristics of both platform concepts with a typical payload complement for each version.

5. FUNCTIONAL PLATFORM MODULARITY

The two platform options offer an interesting distinct approach in their fundamental modularity concept to solve the multimission capability requirements. Option A is based on the philosophy of a generic bus. This means that all necessary functions, even payload dedicated, mission peculiar functions belong physically and functionally to the utility module. These functions must be capable for all missions foreseen with the platform, to benefit as much as possible from the cost advantage of recurring hardware. Some over-design, in particular for the first mission and rigid instrument design requirements are a natural consequence. Option B added between the utility and the payload module a third module, the payload equipment bay, comprising all mission peculiar hardware. Consequently the utility module is a totally recurring item



| | Option A | Option B |
|------------------------------------|---------------------------|---------------------------|
| Instrument Mass (incl. I/F H/W) | 2800 kg | 2000 kg |
| Instrument Power | | |
| Average | 2700 W ¹ | 2600 W |
| Day | 2700 W ¹ | 2600 W |
| Night | 2300 W ¹ | 2600 W |
| Peak | 3200 W | - |
| Pointing Accuracy | < 0.1° (2σ) | < 0.05° (3σ) ² |
| Pointing Stability | | < 0.00065 s ⁻¹ |
| Data Rates | | |
| x-Bd to ground | 4 x 100 Mbps ³ | 2 x 100 Mbps |
| Ka-Bd to DRS | 4 x 100 Mbps ³ | 2 x 100 Mbps |
| Data Record. Cap. | 2 x 30 Gbit | 2 x 30 Gbit |

¹) Design to actual System Requirement;

²) With star trackers;

³) Direct broadcast not included

Tab. 7: Comparison of Payload Relevant Platform Characteristics

and restricted to real fundamental functions. It does not provide payload control functions, payload power distribution, payload data handling and transmission because these functions are mission peculiar, and therefore consequently part of the payload equipment bay. It is important to stress here that the separation of these mission peculiar functions from the basic generic functions is necessarily a physical and functional separation in order to achieve simple interfaces. The payload controller may serve as one typical example to substantiate this functions separation requirement: Separation of the payload control computer only functionally but not physically from the main computer, which is possible also in a generic bus concept, will cause enormous effort in the AIV programme on UM and PM level to simulate the very complex module interface. In case of a physical and functional separation of the master and the payload controller the functional complexity of the interface is very much reduced. Consequently parallel and separate integration of both modules is eased, which is considered mandatory for a system of such complexity. Similar arguments apply to payload data handling and transmission as well as to payload power distribution. Therefore a physical and functional separation of generic utility module functions from mission peculiar functions are seen as an essential requirement for the Columbus Polar Platform.

6. CONCLUSIONS

Based on a detailed mission analysis, the current set of nominal mission and orbit requirements could be confirmed as an acceptable compromise. Both platform concepts are capable to carry the required sets of payload complements and thus demonstrated their basic multimission capability. Both platforms offer enough margin to accommodate the payload complements from a mass and volume point of view, while payload power and data handling are more critical. Further optimization payload on operations seems necessary, in particular if realtime transmission via DRS will be used extensively. Nevertheless the resulting instrument duty

cycles will in fact not challenge the user requirements.

The payload accommodation exercise, performed in parallel to the Polar Platform spacecraft design studies substantiated the flexibility of both concepts, although basic different design features, like S/C flight orientation, resulted in distinct payload module configurations. Sensor and cooler field of view requirements can be met adequately, thermal requirements need more detailed evaluation but could not be seen as critical. The EMC payload requirements can be met as well, given some modifications (frequency change, filtering) can be accepted by the concerned instruments. Flying transversal to the flight vector obviously eased the accommodation of several instruments.

Physical and functional separation of the utility module from the payload module is an essential characteristic of a multimission spacecraft and consequently a requirement to be applied to both platform concepts. This is also confirmed by ERS-1 experience.

7. GLOSSARY

| | |
|-------|---|
| ALT | Radar Altimeter |
| AMRIR | Adv. Medium Resolution Imaging Radiometer |
| AMSU | Advanced Microwave Sounding Unit |
| A.O. | Announcement of Opportunity |
| ARGOS | Data Collection & Localisation System |
| ATLID | Atmospheric LIDAR |
| AURIO | Auroral Imaging Observatory |
| DB | Direct Broadcast |
| DRS | Data Relay Satellite |
| ERBI | Earth Radiation Budget Instrument |
| ERS | ESA Remote Sensing Satellite |
| FPI | Fabry Perot Interferometer |
| GEM | Global Electrodynamics Monitor |
| HRIS | High Resolution Imaging Spectrometer |
| HRTIR | High Resolution Thermal Infrared Radiometer |
| LISA | Limb Sounder for the Atmosphere |
| MERIS | Medium Resolution Imaging Radiometer |
| MIMR | Multifrequency Imaging Microwave Radiometer |
| PAF | Particles and Field |
| PDHT | Payload Data Handling and Transmission |
| PEB | Payload Equipment Bay |
| PM | Payload Module |
| SAR | Synthetic Aperture Radar |
| SARA | Advanced Synthetic Aperture Radar |
| SCATT | Wind Scatterometer |
| SEM | Space Environment Monitor |
| UM | Utility Module |
| VHROI | Very High Resolution Optical Imager |

8. REFERENCES

1. Jenkins, R. et al., Columbus Polar Platform, Option A, C-Zero Final Review Data Package, Bristol, January 1989
2. Cornet, J. et al., Columbus Polar Platform, Option B, C-Zero Final Review Data Package, Toulouse, January 1989
3. Benz, R. et al., Multi-Mission and Programme Aspects Study, Final Report, COL-DO-PPF-RP-0007, Friedrichshafen, April 1989
4. Benz, R. et al., Columbus Instruments Data Sets, COL-DO-PPF-LI-0001, Friedrichshafen, Jan. 1989

UPGRADE TO THE AUSTRALIAN CENTRE FOR REMOTE SENSING A TOTAL GROUND STATION SOLUTION

Jim Friedel
MacDonald Dettwiler
13800 Commerce Parkway
Richmond, B.C., Canada V6V 2J3
Tel: (604)278-3411 • Telex: 04-355599 • Fax: (604)278-0531

Carl McMaster
Australian Centre for Remote Sensing
P.O. Box 28, Belconnen, ACT 2616
Tel: (062)52 4111 • Telex: 61510 • Fax: (062)516326

ABSTRACT

The upgrade to the Australian Centre for Remote Sensing (ACRES) provides total capabilities for the recording and processing of Landsat Thematic Mapper (TM), SPOT Haute Resolution Visible (HRV) and National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) sensor data, as well as user interfaces and production control.

The upgrade is comprised of two sites: a Data Acquisition Facility (DAF) and a Data Processing Facility (DPF). The DAF has the facilities for recording TM, SPOT and AVHRR data as well as for quick processing and transmission of imagery via a microwave link. The DPF has facilities for order entry and processing, production control, image processing and analysis, catalogue update and inquiry and accounting.

The upgrade system produces a wide variety of high quality Computer-Compatible Tape (CCT) and film products ranging from raw to geocoded with subpixel accuracy.

The anticipated effect of the expanded range of products and services in the Australian user community is described.

Another modification organized by the CSIRO and sourced from Australian suppliers has provided a Marine Observation Satellite (MOS-1) reception and minimal data processing capability at ACRES, Alice Springs.

Following a competitive tendering process, a contract was let to MacDonald Dettwiler in July 1987 to provide ACRES with a recording and product processing capability for Landsat TM, SPOT and NOAA Advanced Very High Resolution Radiometer (AVHRR) data. An X-Band reception and Zenith pass capability is to be provided by Datron Systems under a separate contract.

2.0 UPGRADE OVERVIEW

The upgrade is comprised of two sites: the Data Acquisitions Facility (DAF) and the Data Processing Facility (DPF). The upgrade to the DAF included facilities for recording Landsat TM, SPOT Haute Resolution Visible (HRV) and NOAA AVHRR sensor data. In addition, the DAF has capability for performing limited processing at real-time rates and transmission of imagery to the DPF via a microwave link.

The upgrade to the DPF involved a complete product generation and control system and includes facilities for imagery processing and analysis, user order entry, catalogue update and inquiry, production control, and accounting.

1.0 INTRODUCTION

The Australian Government's principal agency for the reception, distribution and processing of satellite remote sensing data is the Australian Centre for Remote Sensing (ACRES). The Centre was originally established as the Australian Landsat Station under the Department of Science & Technology in 1980, but is now part of the Australian Surveying and Land Information Group within the Department of Administrative Services.

ACRES has routinely received LANDSAT Multi Spectral Scanner (MSS) data at its antenna facility in Alice Springs since 1980. This site enables data reception over the total Australian continent and parts of Indonesia and Papua, New Guinea. High Density Digital Tapes (HDDTs) are air-freighted to ACRES Canberra daily, for cataloguing and archiving on the MSS system which is based on Perkin Elmer (Concurrent) processors. The MSS reception facility in Alice Springs and the processing equipment in Canberra were built and installed by MacDonald Dettwiler.

In 1986 a collaborative project between ACRES, the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the Australian Mineral Industries Research Association (AMIRA), established a low-cost temporary modification to the Alice Springs facility enabling the reception of Landsat Thematic Mapper (TM) data and processing to Computer Compatible Tape. This minimally processed product has satisfied the immediate demand for TM data from the experienced users.

2.1 Data Acquisition Facility

This operational facility is situated at Alice Springs, Northern Territories. While having to maintain high system availability, the DAF has a low-key role in terms of overall product generation. Figure 2-1 details the information flow at the DAF. The upgrade to the DAF provided the following major operational capabilities:

- Recording and playback for Landsat TM and SPOT downlink satellite data,
- AVHRR Data Acquisition and Archival System, capable of unattended operation,
- Quick Image Capture System (QICS), for processing of subsampled full scene and full resolution sub-scenes TM, SPOT and AVHRR imagery and remote file transfer to the DPF. Figure 2-1 details the information flow at the DAF.

Raw Landsat TM and SPOT digital image data, together with the station's local time code, are recorded directly onto HDDT. This data can be used as a source of quicklook imagery data at the DAF but is normally transported to the DPF for cataloguing. The image data provided by the AVHRR sensor is stored initially on disk and then archived onto CCT.

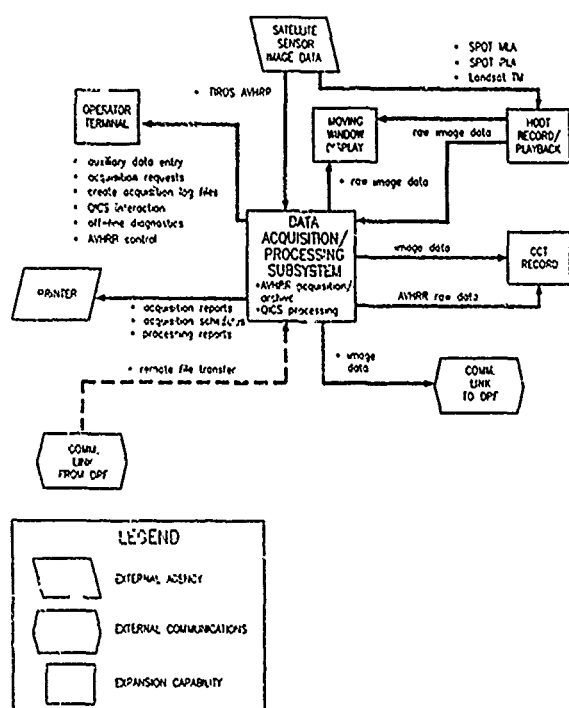


FIGURE 2.1 INFORMATION FLOW FOR THE DATA ACQUISITION FACILITY

To display Landsat TM and SPOT data during real-time acquisition or playback from HDDT, Format Synchronizers and a Moving Window Display (MWD) processor and monitor are provided. The MWD can also be utilized to view AVHRR data on reception, or playback from disk.

The AVHRR Acquisition and Archival System is based on MacDonald Dettwiler's proven Meteorological Data Acquisition System (METDAS). It provides facilities for orbit prediction and reception scheduling in addition to its acquisition and archive role. It also has provision for interfacing directly to an antenna control unit so that unattended data acquisition is possible.

The QICS system provides a mechanism for performing basic geometric and radiometric corrections on subsampled full scene data at real-time rates or on full resolution subscene data at reduced rates. The processing can be done on full passes or single scenes. The processed imagery is then transmitted via a dedicated communications link to the DPF for immediate use, thereby eliminating the time delay caused by having to ship the raw data tapes.

2.2 Data Processing Facility

The DPF is situated at the main ACRES facility in Canberra, ACT. This facility generates all user products and forms the main focus within the user community for enquiries and product generation.

Product generation functions involve all steps commencing at reading the raw data from HDDT (or CCT in the case of AVHRR data), to generating the final product for delivery to the end user.

The DPF can be divided into three major subsystems, each of which has a dedicated CPU. These subsystems are defined as follows:

- **Distribution, Information and Production Control Subsystem (DIPCS)**, which serves as the main external user interface,
- **Interactive and Quicklook Subsystem (IQS)**, which uses the Microimage Quicklook System (MQS) for catalogue and quicklook image production, and a Meridian IAS to perform interactive image analysis functions. The IQS also serves as an interface to the DAF,
- **Image Correction Subsystem (ICS)** which uses the Geocoded Image Connection System (GICS) for the production of bulk, georeferenced and geocoded imagery for TM and SPOT and for the production of bulk and georeferenced imagery for AVHRR. Figure 2-2 shows the information flow at the DPF.

In addition, the subsystems share a MacDonald Dettwiler Color FIRE 240 image recorder for the generation of high quality black-and-white or color film products.

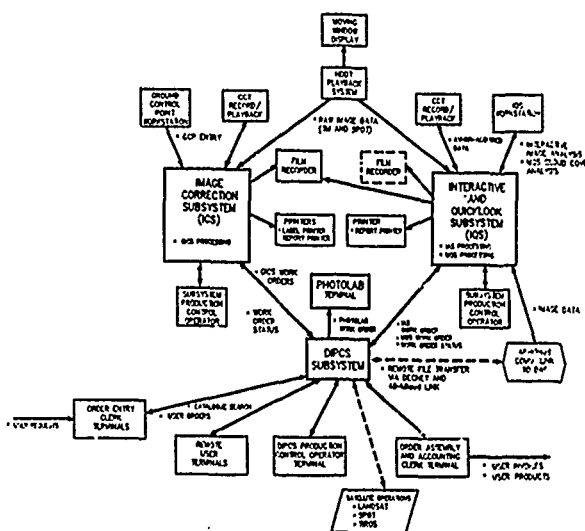


FIGURE 2.2 INFORMATION FLOW FOR THE DATA PROCESSING FACILITY

2.2.1 Distribution Information & Production Control System

DIPCS is an interactive software subsystem which provides the interface between users and the dedicated image processing subsystems. There are facilities for raw and processed data catalogue enquiries and user order entry. Users can enter orders through an order entry clerk or directly themselves, using terminals connected via the Austpac data network.

DIPCS generates work orders for the MQS, GICS and Meridian IAS subsystems, as well as monitoring the status of on-going work on those subsystems. Any film products generated on the subsystems, or requested directly from user orders, give rise to work orders for the photolab. The photolab work orders are handled by DIPCS and the status of on-going photolab work is monitored.

When products are complete, DIPICS passes that information along to the accounting subsystem. Here, packing slips and invoices are generated for distribution to customers. The accounting subsystem also handles other billing and general ledger functions.

2.2.2 MicroImage Quicklook System

MQS, is responsible for generating raw data catalogue updates. When work orders are received from DIPCS, MQS reads and corrects entire passes of TM, SPOT or AVHRR sensor data. Next, interactive cloud cover assessment is performed. Optionally, the system can generate quicklook, daily microfiche and/or cyclic microfiche film products. Catalogue update data is transmitted back to DIPCS.

2.2.3 Geocoded Image Correction System

GICS is the primary product generation subsystem in the DPF for TM, SPOT and AVHRR sensor data products. Product requests are transmitted from DIPCS in the form of work orders. Raw data for the requested products is read, and the necessary correction parameters are generated. If precision products are desired, ground truth is obtained either from previously stored models, from automatic correlation of imagery or using control points marked by an operator.

Using these parameters, GICS resamples imagery to generate raw, bulk or georeferenced products in quadrants or full scenes for TM, SPOT or AVHRR sensors, and geocoded subscene products for TM and SPOT sensors. The products can be resampled into one of many user selectable map projections, using a variety of geoids and resampling kernels. For precision products, an elevation correction can be chosen (a low frequency elevation model containing approximately one point for every 0.25 degree of latitude/longitude) or a full Digital Terrain Model (DTM) correction can be applied. GICS has a pass processing facility for generating models which allow the precision geometric correction of data anywhere in that entire pass with no further ground truth being required.

Once the imagery has been resampled, output products on film or a variety of CCT formats can be generated.

2.2.4 Meridian IAS

The Meridian IAS package, which resides on IQS, is essentially an independent interactive image analysis/enhancement tool. The IAS package contains a complete range of input, analysis and output functions needed to extract available earth surface information from remotely sensed imagery. Work orders transmitted from DIPCS contain instructions for processing imagery that is on disk, obtained directly from GICS, or from CCT. Output products on film or CCT can be generated.

3.0 PRODUCT OVERVIEW

All user products are produced on film or CCTs and may be classed as either 'standard' or 'special'. Standard products are those generated completely within one DPF system while special products may require additional processing by GICS followed by subsequent manipulation by the user before output to film or tape. Film products also require additional development by the photolab. DIPCS is responsible for converting special product requests to work order chains and controlling the progress of the products through the various subsystems.

Film and CCT products produced on any of the subsystems can be recorded in the processed data catalogue once processing has been successfully completed. This allows future products to be ordered as copies by identifying the data from this catalogue.

3.1 Quicklook Products

Quicklook images are corrected only for gross geometric and radiometric errors in the raw data. They are framed according to the standard framing scheme for the satellite/sensor and accompanied by identifying annotation. The quicklook film products are produced at a reduced resolution on black and white, in a 70-mm format on 240-mm roll film (9 per frame).

3.2 Catalogue Products

Catalogue film products are produced in either black and white or colour. Each 240-mm frame of film contains up to two catalogue products, each of which contains eighty-four individual scenes in a 6 x 14 array. The scenes can be organized according to acquisition sequence or, for TM data, by coverage cycle. Catalogue products are generated using imagery obtained from archived CCTs or directly from files generated by data acquisition work orders.

3.3 Standard Products

Standard products are produced on the GICS system and can take the form of either bulk, georeferenced or geocoded data on CCT or high resolution film in colour or black and white. The various characteristics and options for these products are summarized in Table 1-1.

TABLE 1-1 STANDARD PRODUCT CHARACTERISTICS & OPTIONS

| | BULK | GEOREFERENCED | GEOCODED |
|---|------|---------------|----------|
| Geometrically Raw | * | | |
| Resampled to Spacecraft Projection | * | | |
| Resampled to User Selected Map Projection | | * | * |
| Rotated to North Up Orientation | | | * |
| Radiometrically Raw | * | | |
| Radiometrically Corrected | * | * | * |
| Systematic Geometric Accuracy | * | * | * |
| Optional Precision Geometric Accuracy | | * | * |
| Optional Elevation Correction | * | * | * |
| Optional DTM Correction | | * | * |
| Choice of Resampling Kernels | * | * | * |
| Choice of Earth Geoid Models | * | * | * |
| User Defined Framing | | * | |

3.4 Special Products

Products which require some processing on the Meridian IAS are classed as special products. Input data may be from CCT, or from image data residing on disk which has been previously produced by GICS. In addition to generating film or CCT products by applying one of the many image analysis/enhancement tools available, custom format film products (altered scale, resolution, layout, etc.) can be produced using the Meridian IAS.

4.0 POST UPGRADE

4.1 Impact on Operations

ACRES will continue to use the Concurrent system for MSS cataloguing and processing. It is expected that the demand for MSS products will decrease as users begin to use TM and SPOT. However, as the Australian TM and SPOT archive only goes back to 1986, there will always be some requests for pre-1986 MSS.

The longer term options are to maintain and operate the rapidly aging Concurrent system or to shift the MSS cataloguing and processing to the new system. Considerations in making a decision include MSS product demands, Concurrent system performance/maintenance costs,

deterioration of MSS magnetic tape archive, and LANDSAT 5 MSS transmission lifetime.

Some reorganization of ACRES staff has been necessary without significant increase in numbers. Productivity of the system is expected to be high but experience will need to be gained before optimizing procedures. The new DIPCS subsystem which enables on-line catalogue enquiries, product ordering, invoice and statement generation, is expected to be used by ACRES Distributors and regular clients. ACRES User Services staff will be able to spend more of their time servicing the less knowledgeable clients.

4.2 Impact on User Community

A core group of Australian users have a relatively long experience with MSS and other remotely sensed data in a range of applications, especially mineral exploration. These users have well developed methodologies and systems. With the availability of affordable PC-based Image Analysis equipment, many more organizations, including educational institutions, are introducing remote sensing techniques, providing product prices are attractive.

ACRES new georeferenced and geocoded products will reduce the effort needed by the user in the application of the data. Users will need to make the right choices in the sensor type and processing level to best suit their application.

Geographic Information System (GIS) data bases are being implemented by a number of Australian organizations. The availability of geocoded imagery from ACRES can contribute to the economic updating of these data bases.

The Australian Government has recently approved an on-going topographic revision mapping program at medium scales to be undertaken by the Australian Surveying and Land Information Group.

5.0 CONCLUSION

The 1990s will see the launch of new remote sensing satellites and also begin the era of the polar platforms carrying a multiplicity of earth resources sensors.

Data requirement needs will have to be defined and the appropriate processing agreements and capabilities put into place. ACRES in participation with other Australian organizations is well equipped to meet this challenge.

ECUADOR GROUND RECEIVING STATION - COMPONENT SUBSYSTEMS

Ken Cox/Fred Pleasance
MacDonald Dettwiler
13800 Commerce Parkway
Richmond, B.C., Canada V6V 2J3
Tel: (604)278-3411 • Telex: 04-355599 • Fax: (604)278-0531

INTRODUCTION

At any given time, twenty percent of Ecuador is obscured by clouds. Optical satellite data from SPOT and Landsat could only hope to provide part of the national picture. What Ecuador needed for comprehensive coverage was additional access to a sensor providing ground data in all weather by day or night. The MacDonald Dettwiler engineered Ground Station in Quito will provide just that. By late 1989, Ecuador will have reception and processing facilities for Landsat-TM, SPOT and ERS-1 data. This processing range gives Ecuador the unique capability to process data not only from conventional optical satellites but also from a radar satellite with data capture ability denied to sunlight dependant sensors.

Work on the Ecuador Ground Station (EGS) began in 1988. Although MacDonald Dettwiler had worked on other South American remote sensing facilities before, the EGS was the company's first complete South American Ground Station. The system will be installed and operational by September 1989.

The Quito based EGS built for CLIRSEN can be conceptually broken down into three subsystems:

- The Data Acquisition System (DAS);
- The Recording and Playback System (RPS); and
- The Data Processing System (DPS).

DATA ACQUISITION

The DAS, as the ground station front-end, provides interface between the satellite and the RPS. It is designed to ensure a high level of operation reliability and availability so as not to lose more than two percent of available passes due to equipment failure or malfunction. The DAS consists of four major subsystems:

- antenna and antenna control;
- data and tracking receivers;
- test modulator; and
- boresight.

The antenna and antenna control equipment, including associated Radio Frequency (RF) equipment, can receive satellite downlink RF signals in X-Band (8025 to 8400 MHz). The antenna can track Landsat, SPOT and will be able to track ERS-1 and other polar orbiting satellites with similar orbital and downlink signal characteristics. The antenna includes a mechanism to allow tracking of zenith passes without loss of data. The received satellite signals are converted to a 375-MHz intermediate frequency for input to the

receivers. The antenna also provides for the injection of test signals into the RF reception chain, for system loop-back testing.

The data receiver is a single multipurpose unit designed for downlink data capture from polar-orbiting remote-sensing satellites. It consists of a demodulator for Quaternary Phase Shift Keying (QPSK) or Unbalanced Quaternary Phase Shift Keying (UQPSK) data and three bit synchronizer modules. The bit synchronizer modules are designed to recover the original digital data stream with an accompanying clock for a particular satellite.

The tracking receiver is a general purpose telemetry receiver which is configured for operation with the antenna control system to provide X-Band autotracking.

The test modulator is a general purpose unit which generates a modulated RF signal for injection into the antenna. The test modulator output can be modulated with simulated satellite data or pseudorandom bit sequence test signals supplied by the RPS. The test signals are used to measure and verify DAS and RPS equipment.

The remotely located boresight transmitter generates an unmodulated X-Band RF signal used to verify the antenna RF performance and physical orientation.

RECORDING AND PLAYBACK SYSTEM

The RPS consists of two High-Density Digital Recorders (HDDR's), a programmable Data Matrix Switch (DMS), Frame/Format Synchronizers for Landsat-TM, SPOT and ERS-1 SAR.

The RPS includes test data generators for Landsat-TM, SPOT sensor data. These test generators are usable in a variety of system tests including full end-to-end testing of the DAS and RPS. The RPS Bit Error Rate (BER) test set generates and analyses pseudorandom digital data bit stream data to test the antenna, receivers, HDDRs and equal routing equipment.

The RPS provides the facilities used to record satellite downlink data for archiving. Its other functions include formatting data for input into the DPS and providing visual display of image data via a Moving Window Display Processor (MWDP) and video monitor. The RPS also produces test data in the satellite downlink formats and as a Pseudorandom Bit Stream (PRBS).

Data is formatted for input to the DPS by Landsat-TM, SPOT and ERS-1 SAR format/frame synchronizers. Data output from the SPOT and TM format synchronizers can be selected for viewing during downlink data acquisition or tape playback on the MWD. Also viewable on the MWD is data output to the DPS film recorder.

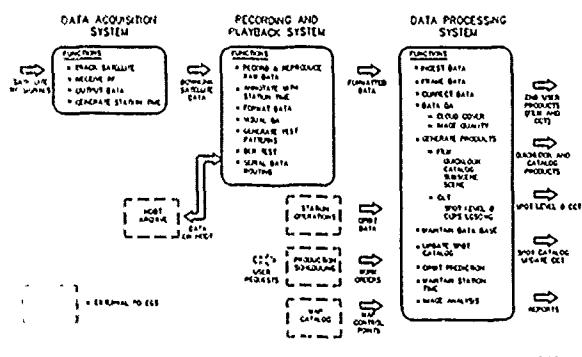


FIGURE 1 FUNCTIONAL OVERVIEW OF THE EGS SYSTEM

DATA PROCESSING SYSTEM

The DPS is based on Digital Equipment Corporation (DEC) VAX-series minicomputers which hosts software packages (described later) to process Landsat-TM, SPOT and ERS-1 SAR data. It provides the functions of data quality assessment and reduced-resolution image product generation for TM and SPOT, and full-resolution processed image product generation for TM, SPOT and ERS-1 SAR. To economically provide a reasonable level of throughput for the large image data volumes of high-resolution satellite image data the DPS employs specialized high-bandwidth processing equipment.

All of quicklook, catalog, SPOT Image, and user products for TM, SPOT, and ERS-1 SAR data are generated in the DPS. DPS functionality essentially falls into the categories of Raw Data Ingest, Product Generation, and Data and Product Quality Assurance. It also provides certain station management functions. The EGS software, written in Fortran and Pascal, runs in the DEC VAX/VMS operating system environment. The software is entirely resident in the DPS and consists of support utility functions and the following four subsystems:

- Microimage Quicklook Subsystem (MQS);
- Geocoded Image Correction Subsystem (GICS);
- SAR Processing Subsystem (SARPS); and
- MERIDIAN (Image Analysis Subsystem).

The Microimage Quicklook Subsystem (MQS).

MQS is used for performing cloud cover assessment and the generation of quicklook and catalog images, and the generation of SPOT catalog update CCTs.

The MQS software consists of code running on the host computer and the Aptec IOC. Software functionality is divided into workstation functions and work order functions. Workstation functions support operator activities to initiate work orders and enter auxiliary data. These functions also provide the operator with system and progress feedback. The work order functions control data acquisition and product formatting and generation. The MQS maintains several data bases. These data bases contain standard configuration and hardware device information, satellite orbital parameter data and standard processing parameter data.

Products generated by MQS are:

- SPOT and TM 70-mm quicklook images in black and white;
- SPOT and TM microfiche catalog images; and
- Reports.

SPOT and TM quicklook products are produced in black-and-white using the FIRE film recorder from a single band of sensor data. The image data is essentially raw. TM scanline reversal and nominal alignment are always applied, and earth rotation and contrast enhancement options are available. SPOT quicklook scene images are produced at a nominal scale of 1:1,500,000. Each scene contains 1000 lines x 1000 pixels of image data plus grey scale annotation. TM quicklook scenes are produced at a nominal scale of 1:4,800,000. Each scene contains 770 lines x 790 pixels of image data plus annotation and grey scale.

SPOT and TM microfiche products can be produced in black and white or in color using the FIRE film recorder. Each 240-mm frame of FIRE film can contain two microfiche products, each of which contains up to 84 individual quicklook scenes. Microfiche scenes for TM are subsampled to 385 lines x 390 pixels of image data while those for SPOT are subsampled to 500 lines x 500 pixels. Each scene has an associated annotation area, enhanced grey scale and geographic tick marks. Microfiches can also be produced in acquisition sequence or by geographic area.

The Geocoded Image Correction Subsystem (GICS).

The GICS provides a precision correction facility for processing imagery from the Landsat-S TM sensor and from the SPOT-1 MLA and PLA sensors. The main imagery input to GICS is raw imagery containing attitude and orbit-related errors, sensor-related errors and scene-related errors. The main function of GICS is to remove geometric and radiometric errors from the imagery and project the corrected images onto standard coordinate systems. The outputs of GICS are film images and CCT products.

The GICS software consists of a number of multitasking processes that run on the host computer, the Aptec Input/Output Computer (IOC) and the Floating Point Systems (FPS) array processor. The main processes which run in the host computer provide workstation support for the operators and process work orders which control the actual data correction functions of GICS. The data input operations and some of the image correction functions are implemented in the Aptec IOC. Image resampling is performed in the FPS array processor.

GICS workstation operators have access to separate production control and control point marking processes. To allow independent activity there is one process per operator. Process isolation further ensures that effects of operator error are restricted to a single work order.

Work orders define the input data for ingest, the corrections to be performed, and the products to be generated. Several work order processes may be active in parallel at different processing stages.

The GICS interfaces with the RPS, the MQS, SPOT Image and with the operating staff. The interface with the RPS is mainly for data input, while the interface with MQS is via shared data bases. The interface with SPOT Image is to provide Level 0 CCTs from GICS. The interface with the operations staff is via the workstations for work order entry and intervention request servicing including interactive control point marking and product data quality assessment.

The GICS produces the following types of imagery products:

- Raw, Systematic or Precision Corrected Scenes and TM Quadrants;
- Systematic or precision Geocoded Subscenes.

Raw, systematic and precision image products are produced as full scenes for SPOT and as full scenes or quadrants for TM. Raw products are similar to bulk products except that geometric corrections (except for TM scan reversal) are omitted. Radiometric corrections may be selected by the operator.

Geocoded subscenes are corrected for earth rotation and other geometric errors.

The imagery is resampled and rotated to align with and correspond to an integral number of UTM map sheets. The pixel sizes for geocoded images from different sensors are in integer ratios allowing these images to be conveniently overlaid in digital form.

For all products two levels of accuracy are available:

- systematic using a priori data, and
- precision using ground truth.

All full scene, TM quadrant and geocoded subscene products can be produced as 240-mm colour film products exposed on the Color FIRE-240 film recorder, and as CCTs in Canada Centre for Remote Sensing (CCRS) LGSOWG format. The SPOT Level 0 CCT products are produced only on CCT in the CRIS SPOT Level 0 format.

A SPOT Level 0 CCT product is generated as an operator-specified work order option for SPOT data processing. This product is used to supply raw SPOT data to SPOT Image.

A Work Order report is printed following the completion of each work order. This report identifies the product(s) generated, the processing options used, and includes Quality Assessment reports for the processed imagery.

The Synthetic Aperture Radar Processing System (SARPS).

The SARPS is an independent subsystem. It does not operate in parallel with other data processing subsystems in the ground station. It interacts with the operator to receive processing parameters and processing commands. Its operation starts with ingesting data from the HDDT. This is followed by data processing and image CCT generation. If required, a film product can be subsequently generated using a DPS utility function.

In terms of hardware, the SARPS uses a subset of the DPS main computer. Data from the ERS-1 High Rate Frame Synchronizer is fed into the Frame Synchronizer Interface (FSI). The FSI unpacks the SAR raw data from 5-bit to 8-bit format and sends it to the Aptec IOC. The Aptec IOC's multiple purposes include providing a high bandwidth mass memory storage for data buffering, allowing the host to communicate with other hardware components and computing statistics on the input data. The input data, stored on a raw data disk, is sufficient to generate a 100 x 100 km image. Input elements on the disk are stored in 8-bit I, 8-bit Q format. The array processor receives control parameters from the host and performs all compute-intensive SAR signal processing operations such as FFT, matched filtering and interpolation.

The MERIDIAN Image Analysis System.

The MERIDIAN subsystem provides an interactive image analysis facility for the post-processing of image data products from the GICS and SARPS, and from external agencies. Its multiprocessing capability allows a number of applications jobs to be run in parallel from a single workstation. MERIDIAN allows the operator to interactively manipulate the image data and perform operations such as classification, feature enhancement, selecting a standard image subset, and mosaicking together two images to form one.

The MERIDIAN software consists of a number of processes which run on single computer system. The MERIDIAN operator is served by an image workstation, which makes use of the hardware of the DPS marking/QA workstation used by GICS and MQS. The software is organized as a set of functions which are executed under operator control via an operator workstation interface. These modules are grouped into two software packages: MERIDIAN Core and MERIDIAN Image.

The MERIDIAN Core package consists of a set of functions and services for controlling the workstation functions, image data base contents, and the software operating environment.

The MERIDIAN Image package provides image processing and analysis functions for activities such as spectral enhancement, classification, statistics, distance measurement, report generation, data input and output, and product generation.

SUMMARY

The MQS and GICS provide data processing functions for Landsat-TM and SPOT, while SARPS provides data processing functions for ERS-1 SAR. The MQS generates quicklook and catalog images, the GICS generates bulk-processed and geocoded image products, and the SARPS generates bulk-processed products. The MERIDIAN system provides interactive image analysis functions for data processed by the GICS and SARPS subsystems.

The GICS, MQS and SARPS are hosted on the main DPS computer system based on a DEC VAX 3600 augmented by high-throughput image data input and processing peripheral equipment incorporating an array processor. Other peripherals and interfaces necessary for satellite data input and product generation are also included in the system.

The MERIDIAN system is hosted on a second VAX 3600 which shares peripherals of the DPS VAX over an Ethernet local area network.

- ARISTOTELES -

Description of the Earth Gravity Field Recovery Mission

R. Benz, M. Langemann, M. Gramolla and G. Mecke¹⁾

Dornier GmbH, Postfach 1390, D-7990 Friedrichshafen, FRG

¹⁾ ESTEC, P.O. Box 299, 2200 AG Noordwijk, The NetherlandsAbstract

Under contract to the European Space Agency a system study for a spaceborne Earth Gravity Field Recovery mission called ARISTOTELES has been performed in 1988, to prove system feasibility covering as further mission objectives geodetic point positioning in the cm range and Earth Magnetic Field measurements (R. Benz et al. 1988). Under the given programmatic constraints, a six months Earth Field Recovery mission in a 200 km near polar dawn dusk orbit has been chosen to recover the gravity field anomalies on a global scale better than 5 mgal, coherently with the magnetic field better than 2 nT with a spatial resolution of 100 x 100 km half wavelength.

The primary payload is a planar gradiometer comprising four accelerometers ultrasensitive in radial and across track direction. A microwave tracking system will provide precise orbit data required for the field restitution and will be used optionally for geodetic point positioning requiring however an orbit raise to at least 700 km altitude and an additional three years mission duration. The payload is optionally completed by a magnetometer and a GPS receiver. The latter one for operational orbit determination and as support of the GRADIO to recover the lower harmonics of the gravity field.

The approval for this optional programme is expected in 1989. Launch is scheduled for the mid 1990's with an ARIANE launcher.

Keywords: Solid Earth Programme, Gravity and Magnetic Earth Field Recovery, Geodetic Point Positioning

1. Introduction

In the mid 1990's, the ARISTOTELES spacecraft will begin to investigate the Earth's fields. The provision of gravity and magnetometry data with unprecedented accuracy and on a global scale will be a milestone for the progress in geodynamic science.

Geography, geodesy, oceanography and other areas will benefit as well as practical applications in satellite navigation, climate modelling and earthquake and volcanism prediction research. Thus the understanding of our planet will greatly be enhanced by the ARISTOTELES mission.

The name of the mission ARISTOTELES stands for "Application and Research Involving Space Techniques Observing the Earth field from a Low Earth Orbiting Satellite" and reminds of the great greek philosopher Aristotle who was the first to speak of gravity forces.

2. Scientific Background

The science of geodynamics is a highly interdisciplinary and complex field that studies the physical forces and processes affecting the Earth's core, mantle and crust and is concerned with interactions of the Solid Earth with its atmosphere, its oceans, with the moon and the sun.

Space techniques are proving to be a powerful tool to study links between individual processes in a short time on a global scale with homogeneous high accuracy and promise to be a major factor in achieving a better understanding of how our planet works. Several scientific objectives are mentioned below. A more detailed background is given by the SESAME report (SESAME, 1986)

GRAVITY FIELD

The structure of the gravity field reflects irregularities in mass distribution in the interior of the Earth. These irregularities are the result of the internal heating and of gravitational attraction on the material assemblage of varying chemical and physical composition. Consequently, knowledge of the Earth's gravity field is the most direct information source for a better understanding of Solid Earth physics.

MAGNETIC FIELD

The Earth magnetic field is made up by the superposition of several components. The temporal variations of the field and the occurrence of magnetic jerks are of primary interest for the scientists. Correlation has been discovered between the westward drift of the field and the length-of-day-variation. There is also a magnetic coupling between the core and the mantle. To fully understand the Earth dynamo effect is one of the greatest remaining challenges in fundamental modern Solid Earth physics.

GEOPHYSICS AND OCEANOGRAPHY

The Earth's surface is made up of individual rigid lithospheric plates which are assumed to float on a relatively soft layer called asthenosphere. These plates are moving relatively to each other with averaged velocities of several centimeters per year, causing earthquakes and volcanism where they collide, separate or slide past each other. The driving mechanisms of plate motion are not yet fully understood, but thermal convection and gravitational forces are believed to be the main contributors. In addition, geodesy and oceanography also benefit in a very profound way from an accurate global knowledge of the gravity field.

Earth Rotation

In response to the gravitational attraction of the sun and the moon, the Earth carries out a rather complicated motion in space. It precesses, the pole of rotation is moving irregularly and superimposed with almost constant drift, the spin rate is slightly changing. Core-mantle processes, earthquakes and fluctuations in the atmosphere are assumed to be the main contributors to these irregularities. Observation of the rotational motion of the Earth by space techniques has resulted in major progress. Further improvements can only come from more precise, more continuous and more dense observations, only achievable with very advanced space techniques.

Earthquake and Volcanism Prediction Research

All three mission objectives of the ARISTOTELES spacecraft shall contribute to the progress in earthquake and volcanism prediction. Detailed gravity information over typical geological features, obtained from the gravity mission, is necessary for the development of accurate earthquake prediction models. The point positioning mission can reveal the relation between major earthquakes and the drift rates of lithospheric plates. Earthquakes are also linked to the Chandler wobble, a component of the polar motion. The understanding of this and other forms of Earth rotation variation will greatly benefit from the magnetometer mission.

3. Applications Background

The ARISTOTELES mission will greatly help man's understanding of the Earth as a system, but there is more to it than the pursuit of pure scientific interest. Advances in geodynamics can directly support work on such diverse fields as e.g. environmental problems, climate modelling and satellite navigation.

CLIMATE MODELLING

Climate models nowadays face increasing interest as a means to assess and predict the disturbing influence of human activities on the sensitive equilibrium of environmental processes. Information gathered from the ARISTOTELES mission can greatly improve the accuracy and reliability of these models in several ways.

A very precise geoid as obtainable from the ARISTOTELES mission will enlighten research in general ocean dynamics and ocean circulation. As a means of storage and transportation of carbondioxide and heat, ocean circulation is a major factor in any climate

model. The ability to absorb carbondioxide in large quantities is an important regulator of the greenhouse effect.

Closely related to the greenhouse effect, the sea level rise predicted by several UNO studies (e.g. Delft Hydraulics, 1987) will endanger large populated areas of coastal land. Space techniques will in combination with tide gauges allow exact monitoring of sea level variations in a worldwide system. It will further the understanding of this phenomenon and the recognition of the involved parameters. It will allow to discover trends in sea level variations and can eventually provide long term forecasts for individual areas. This observation will be simplified by a connection of world wide height systems, as made possible by a precise geoid and spaceborne geodetic point positioning.

SATELLITE NAVIGATION AND GEODESY

A precise gravity field will be the foundation for improved satellite navigation and the basis for exact orbit computation for future and former satellite altimetry missions. The point positioning mission will support this task by allowing to correct geodetic reference frames with information on the movement of lithospheric plates.

Satellite positioning in combination with a precise geoid knowledge provides us with quasi-leveled heights. For developing countries this is the only way to get medium scale topographic height information for mapping, engineering and exploration in a fast and consistent manner, avoiding costly and time consuming terrestrial methods. In the northern hemisphere, the use of the satellite derived geoid is the only way to obtain height systems and height system connections on a continental scale free of systematic errors for science, mapping and engineering.

NATURAL RESOURCES PROSPECTING

An association of plate tectonics with the occurrence of natural resources has recently been recognized. Detailed gravity information of high resolution serves as a first indicator when prospecting Earth resources.

4. Mission Objectives and Requirements

GRAVITY MISSION

The primary mission of the ARISTOTELES spacecraft is the global determination of the gravity potential field of the Earth. The objective is to determine gravity anomalies with

- an accuracy of < 5 mgal
- and a spatial resolution of 100×100 km

The benefit of the ARISTOTELES mission can be seen from a comparison between the amplitudes of various geodynamic phenomena and the envisaged ARISTOTELES system performance as given in Fig. 1.

To achieve the required highly accurate gravity restitution by means of ground processing, the space segment has to provide at least two diagonal gravity tensor components with an accuracy better than 10^{-11} $1/s^2$ which is equivalent to 0.01 E.U.. With the ARISTOTELES sensors T_{yy} (across track) and T_{zz} (radial) can be measured with the required accuracy, while the

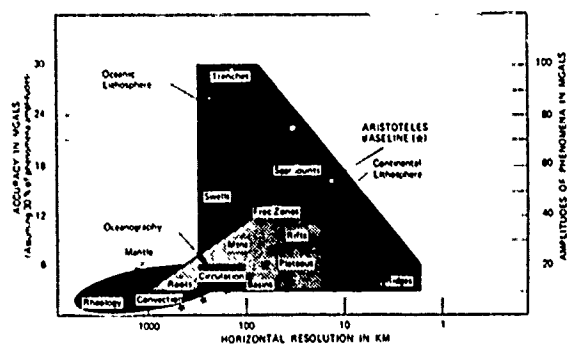


Fig. 1: Comparison of the Amplitude of Geological and Oceanographical Phenomena with envisaged ARISTOTELES System Performance

signal in the flight direction is significantly degraded because of drag forces and consequently less sensitive.

Precise a posteriori orbit restitution is a prerequisite for the accurate gravity field recovery on ground. The required accuracy is about

- < 10 m (radial), and
- < 1000 m (along and across track)

MAGNETOMETER MISSION (OPTIONAL)

Current models could be greatly improved by a detailed global determination of the Earth magnetic field with

- an accuracy of < 2 nT,
- with a spatial resolution of 100 * 100 km.

Orbit altitudes at 200 km and 700 km are of great interest for this optional mission.

POINT POSITIONING MISSION (OPTIONAL)

A precise measurement of absolute and relative positions on the surface of the Earth will enable the correction of geodetic reference frames, the monitoring of Earth rotation and polar motion, of plate motions and plate boundary deformations. The required point positioning precision is

- for baseline up to 100 km: 5 cm,
- for baseline up to 1000 km: 10 cm.

This optional mission requires an orbit raise to at least 700 km and an additional lifetime of 3 years.

5. Mission Scenario

Due to the diverting requirements for the primary and secondary mission objectives a unique mission scenario is required for ARISTOTELES (see Fig. 2). After dual launch together with an ERS-type satellite by an ARIANE into a sunsynchronous, near polar orbit with about 780 km altitude, ARISTOTELES will drift - already in an operational mode of great scientific interest - about 8 months in an orbit with the target inclination to achieve the dawn dusk local time. Then, after a descend phase of one month, the S/C will operate 6 months in the operational orbit for the gravity and magnetic field mission in 200 km altitude. Subsequently an orbit raise beyond 700 km altitude will allow to continue for another 3 years point positioning measurement which could have been started already after in-orbit commissioning of the S/C.

6. Primary Payload

GRADIOMETER

The gradiometer differential measurements of accelerations allow to derive components of the gravity gradient tensor [T]. The instrument named GRADIO comprises four ultrasensitive accelerometer measurements attached to the corners of a quadratic, stiff plate (see Fig. 3). The sensors derive accelerations from the forces that are necessary to keep an electrostatically suspended proof-mass in its nominal position. A calibration device in the instrument center allows to match alignment and scale factors of the individual sensors. The gradiometer provides highly accurate measurements in across track and radial direction.

The instrument is accommodated in the S/C such that its plane is perpendicular to the flight direction and that the instrument and the S/C centers of mass coincide. Isostatic mounting avoids any stress or deformation from being transferred to the instrument.

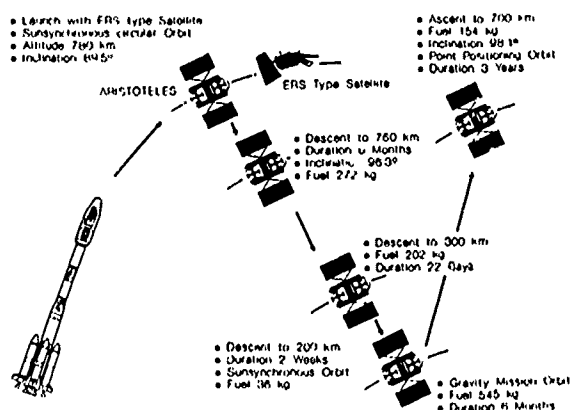


Fig. 2: ARISTOTELES Mission Scenario

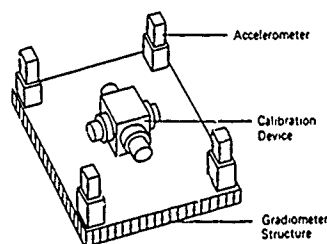


Fig. 3: Gradiometer Configuration

MICROWAVE TRACKING SYSTEM (MTS)

The microwave tracking system has to provide exact orbit determination measurements during the ARISTOTELES mission. The system will comprise a space segment, presumably 7 operational tracking ground stations and a number of scientific ground stations to be defined in the near future. The general tracking principle is outlined in Fig. 4.

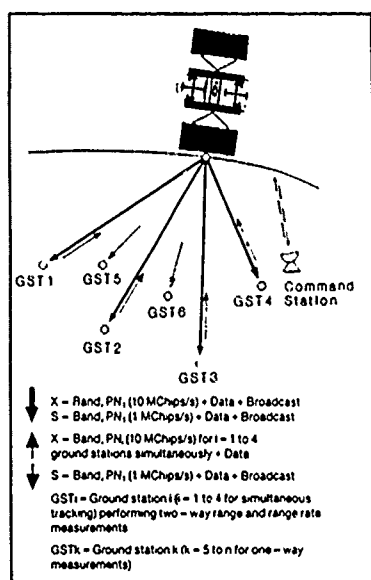


Fig. 4: MTS/PRARE Tracking Principle (Ph.Hartl, 1985)

7. Optional Payload

To benefit as much as possible from the rather unique mission parameters of ARISTOTELES - 6 months in a 200 km, near polar orbit - the payload should be completed by a GPS receiver and a magnetometer in order to support the primary mission and to enlarge the scientific mission objectives.

MAGNETOMETER

Scalar and vector magnetometers have been studied as potential payloads to measure the Earth magnetic field related with core and crustal processes. For the vector magnetometer the EMC requirements together with the stringent pointing knowledge requirement demand for a very stiff boom and an optical alignment system which could not be accommodated under the given constraints. Presently, trade-offs resulted in the use of two already existing fluxgate magnetometers, however only used for scalar measurements as the preferred solution. This allows differential measurements similarly to GRADIO, in order to identify spacecraft impacts on the scientific data.

GLOBAL POSITIONING SYSTEM (GPS)

The use of the well known GPS (see e.g. van Leeuwen, et al. 1985) could ease the operational orbit determination significantly. To meet the orbit determination requirements for the gravity field restitution with GPS seems at least questionable. However the GPS can complete the gravity field recovery in the range of long wavelengths, i.e. in the range where the GRADIO can hardly meet the requirement, due to extreme stability demands.

8. Satellite Configuration

The GRADIO instrument has dictated most of the principal features incorporated in the satellite configuration (see Fig. 5). The need to fly GRADIO at the

minimum possible altitude results in a significant drag force. To provide a mission of sufficient duration for recovery of adequate scientific data, the effect of the drag must be minimised. The cross sectional areas presented to the "airflow" is thus reduced resulting in a long slender satellite body limited by accommodation constraints of the launcher assembly. The electrical energy demand requires solar array wings - in addition to the body mounted solar arrays - but these must be edge on the airflow to minimise drag and large disturbance torques on the satellite and the GRADIO. In consequence it is only possible to fly in near dawn-dusk orbits.

Since the GRADIO instrument measures the gravity field of the satellite as well, unusual consequences result for the satellite internal layout. The instrument must be placed not only at the centre of gravity of the vehicle, but this centre of gravity must also be the neutral point for the gravitational attractions of all components to minimise their influence. Symmetry is thus of major importance, and the heavier units should be placed as far away from GRADIO as possible. The supporting electronic subsystems contained in pods on the sides of the body meet this requirement, while also assisting thermal stability of GRADIO by keeping variable heat generation remote from the instrument.

The largest masses are however the fuel in the 16 tanks. Under manoeuvring accelerations the fuel in the tanks normally moves or "sloshes". To reduce this effect, tanks with stable metal diaphragms are used, restraining the fuel to one end of the tank. The fuel used during the period of GRADIO measurements also is contained only in the outer tanks, i.e. furthest from GRADIO. The fuel in the inner tanks is used during initial manoeuvres to place ARISTOTELES into the correct orbit. In addition, the symmetry requirement can only be met by controlling fuel consumption in the individual tanks, in order to maintain the centre of gravity within fractions of a centimetre of the GRADIO centre. Such a procedure is almost unique to ARISTOTELES.

To achieve the magnetic cleanliness requirements the magnetometers will be accommodated on a deployable boom of about 3 m length, which is stowed along the spacecraft and deployed in the flight direction. This is the most preferred location of the magnetometer to avoid any impact from contamination by thrust plumes and from plasma wake effects.

The MTS antenna will be accommodated at the nadir oriented solar array to guarantee an unobstructed field of view to Earth. Although this is not the best location from a stability point of view, the error contribution is acceptable.

Tab. 1: ARISTOTELES Satellite Characteristics

| | | |
|-------------------------------|--------|----------|
| • Satellite Mass | dry | 960 kg |
| | launch | 1990 kg |
| • GRADIO mass | | 65 kg |
| • Opt. Payload mass | | 35 kg |
| • Power at Solar Array | | 1080 W |
| • GRADIO Power Consumption | | 60 W |
| • Opt. Payload Power Consump. | | 25 W |
| • S-Band Telemetry | | 1 Mbps |
| • Data Storage (36 h) | | > 260 Mb |

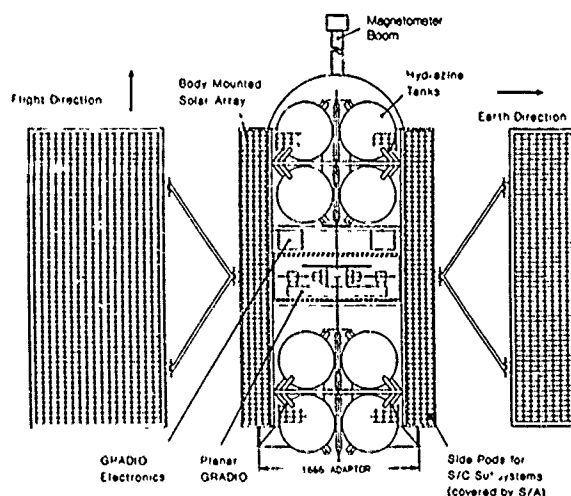


Fig. 5: ARISTOTELES Satellite Side View (+y)

9. Satellite Subsystem Characteristics

The majority of the subsystem for the 3-axis stabilized ARISTOTELES spacecraft are conventional in character with special attention being paid to satellite structural and thermal effects on GRADIO. However, the performance of the attitude and orbit control system is intimately linked with the orientation and control of GRADIO and thus to the quality of the instrument data output. Because the instrument is attached to the spacecraft structure, it is directly affected by drag forces and spacecraft disturbances, placing high demands on control accuracy and stability. To control air drag variations within the bandwidth of GRADIO, aerodynamic flaps are necessary to smooth out the drag variations to one percent of the original values. Consequently particular attention must be paid to the development of a satisfactory AOC/S RCS subsystem.

10. Gravity Recovery System Performance

The gravity field recovery system performance is dictated firstly by the performance of the spaceborne sensor, which is in fact the spacecraft itself and secondly by the performance of the ground processing. Based on a global gravity field recovery simulation (R. Benz, et al., 1988) the basic system requirements could be defined from Fig. 6.

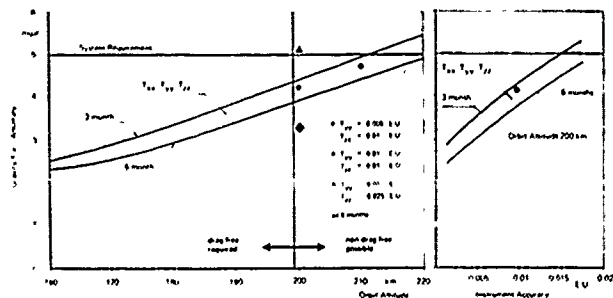


Fig. 6: Relationship between System Performance, Instrument Accuracy and Main System Parameters

The figures demonstrate that in a six months mission the system requirements can be met with sufficient margin by means of a non-drag-free planar gradiometer with 4 accelerometers ultrasensitive at least in two orthogonal directions. The instrument accuracy has to be better than 0.01 E.U. in both directions. Local gravity field recovery could be even better. A detailed performance analysis of the space segment resulted in the gravity measurement performance as outlined in Fig. 7.

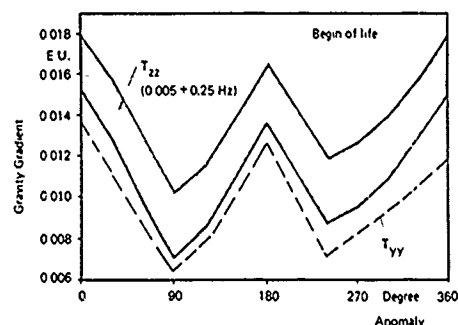


Fig. 7: Gravity Measurement Performance

11. Conclusions

The ARISTOTELES system study (R. Benz et al., 1988) proved that a spaceborne gradiometer can meet an Earth gravity field determination requirement of better than 5 mgal with a spatial resolution of 100 km half wavelength. The required orbit altitude is about 200 km or less with an instrument accuracy of better than 0.01 E.U. given that at least the tensor components T_{yy} (across track) and T_{zz} (radial) were measured.

From a technical point of view the accommodation of a magnetometer and of a GPS receiver is possible in principle as well as the orbit raise for a Point Positioning Mission of additional three years.

Definitely ARISTOTELES could provide substantial information for an improved scientific understanding of the Solid Earth and for various applications of vital human interest.

12. References

1. Benz, R., et al., ARISTOTELES Phase A Study, Final Report, Dornier System GmbH, Friedrichshafen, FRG, 1988
2. Langemann, M., et al., ARISTOTELES Phase A Rider Study, Final Report, Dornier GmbH, Friedrichshafen, FRG, 1989
3. Sesame ESA Special Workshop, ESA SP-1080, Ising, FRG, 1986
4. Study of Impact of Sea Level Raise on Society, Delft Hydraulics, 1987
5. Hartl, Ph., et al., Precise Range and Range Rate Equipment, Inst. für Navigation, University of Stuttgart, FRG, 1985
6. van Leeuwen, A., et al., The Global Positioning System and Its Application, Navigation Journal of the Inst. of Navigation Vol. 26, No. 2, 1979

TWO WAY SATELLITE COMMUNICATION FOR TELEMETROLOGY AND REMOTE CONTROL

H. HANEBREKKE

Division of Telecommunication, The Norwegian Institute of Technology

TRONDHEIM, NORWAY

ABSTRACT – Low data rate satellite communication to fixed and floating buoys at sea, remote observation stations, and fishing vessels is studied. Particular attention is paid to norwegian conditions i.e. high latitude and high mountains. Coverage and reliability measurements have been done along the coast of western and northern Norway, and on major roads in southern Norway utilizing INMARSAT C and PRODAT stations. In the coastal areas we find a reasonable good coverage with only 5% loss of messages when both the AOR and IOR satellites are used from the same location, whereas the land mobile experiments gave 40 to 70 % loss depending on the elevation angle. Presently we are investigating the possibility of using INMARSAT C or PRODAT stations in the major fishing areas between Norway, Greenland, and Svalbard and in the Barents Sea. A new method of data collection from ocean areas based on the fishing fleet is proposed.

Satellite communication – Telemetry – Remote control.

1. INTRODUCTION

The use of satellites has to a large extent contributed to the collection of data from rural areas. The vast majority of data stems from satellites in low polar orbits. These platforms are usually carrying optical or infrared sensors aimed to study particular items on a global basis such as land, ocean, ice cover or atmosphere. Next generation satellites will also carry synthetic aperture radars (SAR) which make it possible to observe land and ocean when covered with cloudes These methods collect a vast amount of data that have to be post-processed. This is a costly procedure which in most cases results in time delay between measurement and presentation. The resolution is often too low to observe phenomena of interest and the observed signature is solely determined by the radiating surface and the underlaying area close to the surface. Thus in many situations it is desirable to perform measurements in situ to obtain more exact information about temperature, chemical composition, and biological contents. The data can be used as stand alone data or in combination with other remote sensing information.

Deployed radio transmitters on buoys at sea and on icebergs have been used for several years. In later years low orbit satellites have been utilized to collect data from such transmitters. Several norwegian companies and laboratories have

used low orbit satellites to convey data from buoys to their base stations in Norway. The Norwegian Hydrotechnical Laboratory (NHL) has used the ARGOS system which is implemented on two NOAA satellites, to collect data from the Arctic Ocean to study ice drift and ice conditions. The same system has also been used to study ocean waves at the coast of Tonga in the Pacific in planning a wave power plant. The Oceanographic Company of Norway (OCEANOR) are conveying oceanographic measurements via the same satellite system from buoys in ocean areas around the earth.

The data collection is relying on one way communication from the buoys. The sensors are preprogrammed to take measurements at regular intervals and transmit the result when the satellite is passing. The data are received from the base earth station after a delay of 3 to 4 hours. There is no possibility to change this procedure when the buoys are deployed. This excludes interactive choice of sensors and remote control of the station if interesting, rapidly developing phenomena occur. It also precludes the possibility to load new software into the station when that is desired. To overcome these inconveniences there is a growing demand for low data rate two way radio links via satellite.

This recognition has led us to undertake a study of communication possibilities for such services with emphasis on

norwegian conditions. Particular attention has been paid to the communication to buoys, drifting icebergs, and the norwegian fishing fleet.

2. DEMAND FOR LOW DATA RATE COMMUNICATION

Mobile satellite communication provides service independent of population density and degree of remoteness and thus have the capability of providing flexible services to thinly populated countries. It also provides a means of unified communication for mobile units which are regularly operated both within and outside the coverage area of terrestrial mobile systems. The Norwegian Teleadministration is this year introducing a mobile data service, Mobitex, in the Trondheim region. By mid 1990 the service will also be available in the regions around Oslo, Bergen, and Stavanger, but it will take several years before the service will be available in the rural areas of Norway. The mobile satellite services seem therefore to be the most reasonable alternative for low data rate communication to these regions in the 90-ties.

The interest for collecting marine data by using fixed or drifting buoys and icebergs has increased markedly the later years. This is mainly connected to the exploitation of oil in ocean areas around the earth, but also by the increase in mapping and monitoring pollution and biological contents in the sea. Modern buoys collect a variety of data and the transfer capacity in existing communication systems is too low. Although a lot of signal processing is done in the buoys the transmission rate is the bottleneck. The use of mobile satellite stations like INMARSAT C or PRODAT seems by first sight to be the solution. However, the power consumption in these stations is 10 to 15 W in idle condition and this is far too high for battery powered buoys. New designs and new methods should be considered to reduce the power consumption when the stations are idle. But this will take time and it is most likely that buoys will have to rely on low orbit satellites for several years.

Perhaps one of the most interesting areas for low data rate mobile satellite communication is the services to the fishing fleet. These include data net service to the vessels, improved security in distress and rescue operations, and telemetry of marine parameters through already available data on board.

Data bases are presently being set up to serve the norwegian fishing fleet and the norwegian fish farming industry. The aim is to collect relevant data from sources in Norway and abroad and make them available for the users via mobile terminals. The system will include such services as fish reports, prices and market conditions for fish internationally, accounting and banking operations, and weather and wave forecast in the actual ocean areas. These services will be

available on request and may be of great importance for the users. Particularly, each fishing vessels can via its mobile station ask for an updated weather and wave forecast for its position whenever needed. Only low data rate communication both ways is needed for these tasks.

Here we propose another mode of operation, namely that of data collection from remote fishing vessels. Most modern fishing vessels are well equipped with electronic measuring instrumentation. Position, wave, wind, current, salinity and temperature may be measured regularly and stored on data bases on board. Even sea temperatures and other quantities below the surface can be measured on the larger trawlers. Through a polling procedure to all the participating ships the information may be read into central data bases ashore without any attendance of the ships crew.

The participation in this system should of course be agreed on by the ships on an individual basis. There may be hesitation in providing the ships position in such a polling system. However, by doing that the position will be registered for each ship and in case of rescue the position is known to the rescuer.

INMARSAT C or PRODAT stations are the obvious choice for this service if coverage and reliability are satisfactory in major fishing areas between Norway, Greenland and Svalbard, and in the Barents Sea. An estimate of 800 to 1000 stations for the norwegian fishing fleet is envisioned.

The norwegian lighthouse system consists of a large number of buoys and lighthouse stations. In the future most of these stations will be unmanned and the stations will be operated either automatically with frequent inspections or remotely controled and monitored. Remote control and monitoring is obviously the most reliable way because operation faults can be discovered immediately and measures be taken to correct the fault. Only low data rates are required. This service could be operated by stations similar to INMARSAT C or equivalent, but should probably not be operated in the frequency bands for mobile satellite communication. Though, if the capacity of the mobile systems are sufficient the inclusion of such services would increase the demand for stations and certainly contribute to reduce the cost of the stations.

The management of hydro power resources is an area where small transportable stations would be valuable. Remote stations for monitoring snow cover and reservoir filling, and remote control of dams are tasks that can be well managed with low data rate satellite communication. Surveillance of remote military storages and pollution monitoring are others. Some of these tasks can be accomplished by fixed or semi-fixed stations. However, the use of mobile stations greatly enhances the flexibility and the station cost will be

lower than for fixed stations. The high power consumption is a problem and remote stations on land must have electric power generators to charge the batteries regularly.

3. SATELLITE SYSTEMS.

The success of implementation of satellite services to mobile stations depends on the reliability of the services. The majority of mobile services that will be offered in the next 5 to 10 years will be based on geostationary satellites with some capacity available in low orbit platforms. The low elevation angle in ocean areas outside northern Norway and the combination of low elevation angles and high mountains in Norway pose particular problems for communication via geostationary satellites. In figure 1 we have shown contours for 0° and 5° elevation angles from the MARECS B2 and INTELSAT MCS-A satellites which are used for the INMARSAT C service together with 0° and 5° contours for a geostationary satellite at 10° E as proposed in the PRODAT concept. Earlier measurements [1], [2] with INMARSAT C and PRODAT suggest that it should be possible in open sea to operate close to 0° elevation with only some reduction in reliability. If this is

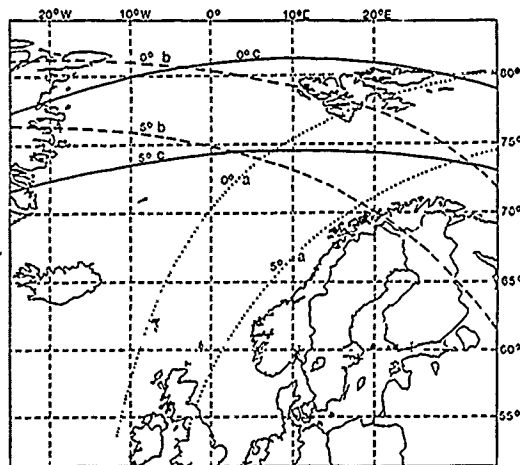


Fig. 1. INMARSAT C and PRODAT elevation angles in the Norwegian Sea and the Barents Sea. a) IOR satellite, b) AOR satellite and c) satellite at 10° E.

correct, the major ocean areas of the Norwegian Sea and the Barents Sea will be covered by geostationary satellites. In coastal waters, in fjords and valleys in Norway large areas lay in shadow from high mountains and can not be reached. The Arctic region is another area which can not be served by geostationary satellites.

A recent study of communication possibilities and demands for the polar regions [3] concluded that the

communication needs in these areas had to be based on low orbit polar platforms for the next 5 to 7 years. The introduction of high elliptical orbit satellites, HEO, for instance in connection with NAVSAT, should then offer continuous service towards the end of the 90-ties.

HEO satellites would solve many of the problems connected to mobile services in Norway and in the Arctic region offering high elevation angles for the entire area. However, no decision has so far been made to implement such a system and communication in the next 5 to 10 years must be based on geostationary or low orbit satellites.

4. EVALUATION OF SERVICE COVERAGE AND RELIABILITY.

The collection of data from buoys via low orbit satellites have proved to be successful. As mentioned above Norwegian companies and research establishments have utilized the ARGOS system to gather oceanographic data from oceans around the earth. The very low data rate capability and the fact that this system provides only one way communication have initiated a search for a more powerful two way communication. New systems such as INMARSAT C and PRODAT are presently being evaluated for this and other purposes. The evaluation is based on field trials.

In 1987 coverage tests of the INMARSAT C system along the coast of western and northern Norway, and inland southern Norway were performed in collaboration with The Norwegian Tele Administration [1]. The experiments were based on the Enhanced Group Call service provided by INMARSAT for coverage tests. Figure 2 shows a sketch of the routes that were followed. Signals from both the Atlantic Ocean, AOR, and the Indian Ocean, IOR, satellites were utilized for the measurements along the coast while the AOR satellite was utilized for the land mobile experiment.

With the IOR satellite INTELSAT MCS-A at 63° E the elevation angle varied between 5.5° and 8.1° along the coast from Bergen to Kirkenes. This is a reasonable high elevation angle and would give good reception in open sea. Due to mountains between the ship and the satellite on parts of the route 33 % of the messages were lost. A closer look at the results shows very good reception in open areas.

With the AOR satellite MARECS B2 at 26° W the elevation angle (corrected for satellite inclination) varied from 18.6° to -0.2° . In this case 16 % of the test messages were lost. All losses were observed in the area north of Lofoten and only one of the received messages was corrupted.

Combining the results from the two satellites only 5 % of the messages were lost from both satellites at the same time.

The landmobile experiments gave an average of 40 to 70 %

loss with elevation angles varying between 17° and 70° . However, long periods were observed when communication was not possible due to mountains, hills, and trees between the mobile and the satellite.

This year a land mobile experiment was done with the PRODAT system. A PRODAT station was mounted on a car and a route from Trondheim to Nordfjord on the west coast and back to Trondheim were chosen as shown in figure 3. This route is a combination of open areas and narrow fjords. The station was operated continuously with the car moving. Telex messages addressed to ourselves were sent to the Villafranca earth station and returned back to the car. Excellent transmission and reception were observed when the station was synchronized to the satellite signal, but for about 50 % of the

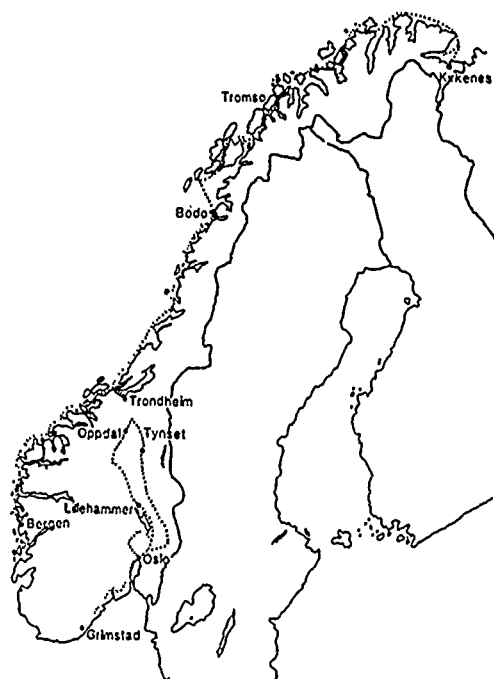


Fig. 2. Test routes for evaluation of INMARSAT C system in Norway.

time the station was out of synchronism because of mountains.

Presently we are investigating the ocean areas between northern Norway and Svalbard. The aim is to test the reliability of INMARSAT C and PRODAT at very low elevation angles and to demonstrate low data rate services to the norwegian fishing fleet. We plan to operate both stations on the same ship for comparison. However, there has been a delay in the delivery of the INMARSAT station so the first cruise will have a PRODAT station only. The stations will be operated on

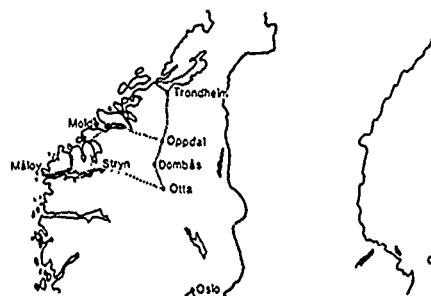


Fig. 3. Test route for evaluation of PRODAT land mobile system in Norway.

a continuous basis by recording information about the received data vectors, position, and weather and wave conditions. Telex service will be tested regularly by sending messages between the ship and the base ashore and by addressing messages back to the transmitter. Provisions have also been made to supply news from a local area newspaper and to transmit meteorologic data for the ships position on request from the crew.

5. CONCLUSIONS

Two way low speed satellite communication for remote data collection and remote control have been studied. The new mobile services like INMARSAT C and PRODAT are promising for many of the tasks. The power consumption, however, is too high for use on buoys and other stations without electric power generation. Tests have been performed to evaluate these stations for use inland and along the coast of Norway. A new system for marine data collection based on the fishing fleet is proposed and we are presently doing field tests on fishing vessels in the ocean between mainland Norway and Svalbard to evaluate the reliability with elevation angles close to 0° .

ACKNOWLEDGMENT

The author wishes to thank S. Berg of The Norwegian Tele Administration for doing some of the field measurements.

REFERENCES

1. Berg, S. A., "INMARSAT Standard-C System", Siv. Ing. Thesis. 46 pages, in norwegian. December 1987.
2. Rogard, R., Jongejans, A., Loisy, C., "Mobile Communications by Satellite. Results of Field Trials Conducted in Europe with the PRODAT System", to be presented at The 8th International Conference on Digital Satellite Communication. Guadalupe, Mexico, 1989.
3. Rørholt, L., "Market Survey and Mission Definition Study on Polar Communications", Telox Engineering Report TEL-124/86. 1986.

INFLUENCE OF THE RELIEF ON RADIOMETRY AND RESOLUTION OF A SAR IMAGE.

Cl. Jung

D. Vidal Madjar

Ph. Rebillard

CNET/CRPE
38-40, rue du Gl Leclerc
92131 Issy-Les-Moulineaux (France)
(1) 45.29.55.39

Sepimage
52, quai de Dion Bouton
92806 Puteaux (France)

Specific geometry of SAR acquisition is highly sensitive to slopes and involves geometric and radiometric deformations on radar imagery. In order to be able to properly interpret informations in mountainous regions, it is necessary to carefully quantify the effects of the relief on the SAR image.

The purpose of this study is to point out the influence of the relief on spatial resolution, in order to correct for radiometric distortions.

The method which is used consists in simulating the radar point spread function and studying how it is distorted by various reliefs.

A preliminary work assimilated the relief to its locally tangent plane and studied the influence on the spatial resolution, of changing, for the tangent plane, inclination and orientation relative to the radar beam.

The logical continuation of this work is to improve to the second order the model of the relief, and take into account the radius of curvature of the relief, especially in the case of very extended radar responses (i.e. when slopes are perpendicular to the beam).

The question is to find a criterion on the radius of curvature and on the slope, according to which the approximation of the relief to its tangent plane is satisfying or not.

For that purpose, we first compute, at each point M of the ground, the radius of curvature of the relief, in the two principal directions of the considered digital elevation model. Then, in those two directions, we assimilate the relief to the osculating circle at the given point M and therefore, we are able to calculate the surface of the point spread function projected on the ground, in order to compare it with the tangent plane approximation results.

The method is tested on a simulated sinusoidal relief, and then the obtained criterion is applied to a real DEM on the mountainous region of the Jura (France).

The final aim of this study is to obtain an analytical computing tool allowing the determination of ambiguous points in a SAR image and the actual losses of contrast and radiometry due to the relief.

MULTIPLE-IMAGE SAR SHAPE FROM SHADING

J. Thomas, W. Kober, F. Leberl
 VEXCEL Corp.
 2905 Wilderness Pl.
 Boulder, CO 80301, USA
 Tel: 303/444-0094
 FAX: 303/444-0470

ABSTRACT

The purpose of the present effort is to explore the combination of techniques for shape from shading with stereo-radargrammetry to produce terrain surface models using multiple SAR imagery.

The use of local variations in pixel shading as an indicator of terrain slope changes represents an opportunity for increasing the accuracy of terrain mapping over that which is available from stereo-radargrammetry alone. Shape from shading can potentially provide a relative changes in height at each pixel. This leads to a denser set of height measurements and a more faithful rendition of the local terrain shapes. However, some additional type of boundary values or terrain low-frequency information is required. These can be obtained from stereo or from altimeter measurements.

Two essentially different types of approaches are considered. Preliminary results using real SAR imagery are encouraging.

1. INTRODUCTION

Shape from shading requires the formulation of the reflectance map, which is an analytic description of the relationship between radiometric image values $I(x,y)$, surface coordinates, and image sensing vectors. We will assume the following parameters are the determinants of the reflectance map, R [Frankot, Chellapa, 87]:

$$(1) I(x,y) = R(Z_x, Z_y, b, l, n)$$

where:

Z_x, Z_y = slopes in x and y directions
 b = illumination vector
 l = boresight vector
 n = albedo

Note that for radar, $b = l$

The image gray values are then the discretized levels between the minimum and maximum radiometric values.

An exact functional relationship among these variables is intractable to derive analytically, and in the past has been the subject of empirical studies and some simplified modeling of the scattering mechanisms involved.

However, in many shape from shading studies, for example [Wildey, 86], the reflectance map is simply the product of a constant scattering cross section coefficient, dependent only on an assumed constant albedo value, and the cosine of the local incidence angle, i.e. the angle between the illumination vector and the local terrain surface normal vector. In the single image case, there is a fundamental ambiguity between changes in albedo and slope when a change in shading is encountered.

Such a model assumes that the scattering cross section coefficient so, does itself not vary with local incidence angle. In order to simplify the computations involved, this is the model that has been assumed in the present study. More complex models of the reflectance map will be used in follow-on efforts.

A major difficulty with single image approaches to shape from shading is the assumption of uniform albedo. Another difficulty with shape from shading is that the problem is mathematically underdetermined without boundary conditions. In general, the shading information from single pixels merely implies a cone constraint for the local surface normal. Continuity assumptions connect the estimated directions for neighboring pixels, and prevent independent solutions for each separate pixel.

Another constraint involves the notion of "integrability". Analytically, this condition implies that:

$$(2) Z_{xy}(x,y) = Z_{yx}(x,y)$$

where $z(x,y)$ is the analytical expression for the height.

This analytic condition is just the intuitive condition that heights can be integrated along any path, since these values are independent of the path of integration. In practice, enforcing this condition also acts as a smoothing process on the computed terrain surface.

Still another constraint is "regularization". This term effectively limits the amount of allowable oscillation in the computed terrain surface.

Two essentially different approaches for reconstructing the surface are discussed in this paper. Approach #1 consists of two variations which involve a multiple image generalization of the method in [Frankot, Chellapa, 87]. The latter approach formulates the reconstruction problem as the minimization of a cost function which contains two types of terms.

The first type of term is a measure of the difference between the pixel gray values of an actual image versus the values predicted using the current estimated terrain model. Because of the mathematically ill-posed nature of such inverse problems [Baltes,80], solutions which minimize this term will lead to very oscillatory terrain estimates.

Therefore, the second type of term is a regularization term which acts as a penalty function to limit the amount of terrain oscillations. In this way, solutions to the combined metric represent a compromise between faithful image prediction and terrain slope variations. The reconstruction problem is thus formulated as a calculus of variations problem whose solution is obtained using the Euler-Lagrange equations.

Such methods tend to be robust with respect to noise because they tend to distribute the effects of noise, rather than accumulating them as in an integration algorithm.

However, most of the earlier work done using the variational approach tended to produce solutions that did not satisfy the condition of integrability. One approach [Horn,86] used a penalty function to drive the iterated solution toward integrable solutions. In practice, this meant that the computed solution usually was only "close" to being integrable. A step in the right direction was taken in [Frankot,Chellapa,87] wherein the latest estimated solution was projected onto an integrable subspace of solutions on every iteration.

However, both variations of Approach #1 require the mutual registration of multiple images. Because such registrations generally involve residual errors, an iteration between registration and terrain reconstruction is generally required.

Approach #2 for multiple imagery involves a set of simultaneous equations which relate slope-induced shading and height measurements to the Fourier-series coefficients of the terrain model.

2. ALGORITHMS FOR SHAPE RECONSTRUCTION

2.1 General

Section 2.2 discusses the two variations of approach #1, the N-image generalization of the method in [Frankot, Chellapa,87]. A very brief description of approach #2 is given in section 2.3.

2.2 Approach #1: Generalizations of Frankot's and Chellapa's Method

The two generalizations of Frankot's and Chellapa's method involve generalizing the cost function or combining separately derived models using projection in Fourier space. These approaches follow in the next two sections. A discussion of the incorporation of spot

heights derived from stereo into the reconstruction is described in section 2.2.3.

2.2.1 Variation A: Generalized Cost Function

One approach to a N image generalization of the approach in [Frankot, Chellapa,87] is to generalize the cost function.

The cost function that was minimized was:

$$C = \int (I-R)^2 + \lambda(U_x^2 + 2U_{xy}^2 + U_y^2) drdy \quad (3)$$

where

I = Actual image

R = Reflectance map

ur = Slope in the range direction

uy = Slope in the azimuth direction

urr = Second partial derivative of u in the range direction

ury = Partial derivative of u with respect to r,y directions

uyy = Second partial derivative of u in the y direction

λ = Regularization parameter that controls the amount of curvature in the derived height model.

A generalization of the algorithm to N images minimizes the following cost function (superscripts refer to image number):

(4)

$$C = \int (I^1-R^1)^2 + (I^2-R^2)^2 + \dots + (I^N-R^N)^2 + \lambda(U_x^2 + 2U_{xy}^2 + U_y^2)$$

Using the generalized cost function for N images we have [Thomas et al,89]:

$$\begin{aligned} U_x &= \overline{U_x} + \frac{3\epsilon^2}{10\lambda} ((I^1-R^1)R_{ux}^1 + \dots + (I^N-R^N)R_{ux}^N) \\ U_y &= \overline{U_y} + \frac{3\epsilon^2}{10\lambda} ((I^1-R^1)R_{uy}^1 + \dots + (I^N-R^N)R_{uy}^N) \end{aligned} \quad (5)$$

In the generalization presented above, it is assumed that the reconstructed terrain model and all images are in the same coordinate system. In general, this assumption is not valid, unless the images are rectified and registered. Because of registration problems, this algorithm is not as effective as the one discussed in the next section.

This model containing the surface facet slopes for each terrain element in the estimated model is converted to a model containing height values using the Fourier projection technique of [Frankot, Chellapa, 87]. This process is performed on every iteration of the complete algorithm. This Fourier-based technique is generalized in the next section.

2.2.2 Variation B: Combining Models Using Fourier Projection

Another approach for generalizing the reconstruction process to N images is to combine single image-derived models in the Fourier projection stage of the algorithm in [Frankot, Chellapa, 87]. This approach is superior to the previous cost function generalization and is discussed below.

In order to handle the case where the different images are in different coordinate systems (e.g. different slant-range projections), an integrability projection based algorithm has been developed. In this approach the slope information (as opposed to pixel values) for each pixel from the different images are combined in the integrability projection stage to derive a composite elevation model. The differences in the co-ordinate systems from one image to the next are handled by a rotational co-ordinate transformation and resampling. The rotational transformation that would be used is exactly the transformation used in [Frankot, Chellapa,87] to approximate the SAR imaging geometry. The following equations are developed for the two image case and are readily extended to the N image case.

In this approach the iterative form for a single image is used to derive a set of slopes for each of the images separately. These slopes are then simultaneously projected on to an integrable surface as follows:

The estimated non-integrable slopes for image (1) are Z_x^1, Z_y^1 , and the estimated, non-integrable slopes for image (2) are Z_x^2, Z_y^2 .

Let the integrable slopes be Z_x, Z_y .

The projection at each iterate is attained by minimizing the following distance function:

(6)

$$d = \int (Z_x - \hat{Z}_x)^2 + (Z_x - \hat{Z}_x^1)^2 + (Z_x - \hat{Z}_x^2)^2 + (Z_y - \hat{Z}_y)^2 + (Z_y - \hat{Z}_y^1)^2 + (Z_y - \hat{Z}_y^2)^2$$

to get the following equation as shown in [Thomas et al. 89].

$$(7) \quad C = \frac{(C1+C2)P_X + (C2+C4)P_Y}{2(P_X + P_Y)}$$

where

$C, C1, C2, C3, C4$ represent the Fourier coefficients of $Z, Z_x^1, Z_y^1, Z_x^2, Z_y^2$ and Z represents the desired integrable surface.

Our experience on synthetic test imagery indicates that convergence was attained with far fewer iterations using the multiple image algorithms than with the single image version.

2.2.3 Reinforcement of Spot Heights

Another modification of the single-image algorithm is the correction of estimated heights in the Fourier projection step described above, to take into account a priori knowledge of some spot heights gained from stereo. Such spot height values represent constraints on the estimated values. Conceivably, such constraints could be reinforced at each iteration.

However, reinforcement at each iteration of a "hard" constraint, such as these spot heights, tends to cause an oscillatory "ringing" in the estimated surface. Therefore, such corrections are presently made in a

"soft" mode. This means that corrections are made using a given fraction of the difference between the constraint value and the estimated value at each iteration. As the iteration number increases, the fraction of the difference between the constraint value and the estimated value can be allowed to increase.

The use of a soft version of a constraint rather than a "hard" constraint is generally useful in iterative optimization problems, since there may be no connected path from a suboptimal solution to the optimal solution which satisfies the hard constraint. However, by continuity, a path exists consisting of suboptimal solutions which satisfy soft constraints.

There were some robustness problems observed in this spot height reinforcement process during the simulations. Near peaks and troughs of a smoothly varying surface, there was an observed "undershoot" and "overshoot" of the constructed surface. The reconstruction process appears to be somewhat ill-conditioned in those regions, since there is almost no difference between predicted and sensed gray-values. The reinforcement, on any iteration, of spot heights in such regions appears to be reduced by the algorithm on subsequent iterations.

Future follow-on efforts will investigate other approaches for reinforcing spot heights within the framework of this algorithm. One possible approach consists of reinforcing entire subregions

around the spot height.

2.3 Approach #2. Simultaneous Equation Representation

An entirely different approach to terrain reconstruction involves posing the problem as one of estimating an integrable solution directly, rather than projecting estimated solutions onto subspaces of integrable functions as was done in the previous algorithms.

This method relies on formulating the reconstruction problem as a problem of estimating the coefficients of integrable basis functions. In this case, the basis functions are chosen to be 2-D Fourier Series basis functions.

Using these basis functions, any existing height constraints can be formulated as linear equations for the basis function coefficients. The predicted image information is formulated as a set of non-linear equations involving the derivatives of the terrain.

An advantage over the previous procedure is that there is a more natural representation of both slope and height information, rather than making height corrections to an algorithm dominated mostly by slope information.

This approach also has the advantage that image registration is not required because it is the terrain Fourier coefficients that are being estimated, not individual terrain cells.

A disadvantage is that a system of nonlinear equations must be solved. One approach is to linearize these equations and use the matrix generalized inverse to minimize the length of the residual vector. The use of the Householder transformation will allow more efficient and stable computation of this inverse.

The cost function in this case is the least-squares norm of the residual vector. The regularization is applied via directly band-limiting the spatial frequencies of the terrain rather than the use of penalty functions.

This procedure has not been implemented, and is still in the conceptual stage.

3. RESULTS

The shape-from-shading algorithm (approach#1 version a) described above was implemented and a number of experiments were conducted to evaluate the performance. The experiments were initially conducted with simulated radar images and later with actual radar images.

Figure 1 is a perspective view of a height model of the Brazzeau range in Canada. The variation in height from the lowest point to the highest is approximately 2000m, the spacing of sample points was 60m, and the size of the height model was 128x128 elements. Using this height model as input, a set of three noise-free radar im-

ages were simulated with the radar assumed to be flying at a height of 1000km and imaging with incidence angles of 65.38, 56.44, and 50.28 degrees to the near edge of the scene. Figure 2 shows one of these simulated images.

A three image version of the same shape-from-shading algorithm was run using the three radar images simulated above. Using a heavily smoothed version of the original height model as a starting condition, and by locally enforcing 225 regularly spaced height points, the algorithm was able to generate a height model. On comparing the derived height model with the actual height model (shown in figure 1), it was found that the standard deviation of the error was 80m. This is shown in figure 3.

The algorithm was then run using a single image as input and using the same initial starting condition (i.e. the same smoothed version of the height model). Figure 4 shows a perspective view of the height model thus derived, and the standard deviation of error on comparing with the true height was found to be 119m.

In order to test the performance of the algorithm with real radar images, a two image version (since only two overlapping images were available at the time) of the same algorithm was implemented. In this experiment, stereo radar images were used to create a height model, and the height model was then used to rectify the images. The rectified and co-registered images were averaged from 6m pixels to 24m pixels. Figures 5 and 6 show the 128x128 pixel input images. The algorithm was also provided the stereo derived height model as a starting model. Figure 7 shows the stereo derived height model as a radar power map (excluding the geometric distortions) created from the same position the true radar created one of the input images (from the left of the image and imaging to the right). Figure 8 shows the output height model derived from the shape-from-shading algorithm presented as a radar power map. Spot heights were not enforced in this experiment, and due to the lack of map data we were unable to compute a quantitative error.

4. DISCUSSION

As seen in figure 8, the multiple image shape-from-shading is capable of providing a height model that appears to be an improvement from the results provided by radar stereo. Although a map derived height model was unavailable for comparison, the final height model seems to be consistent with the input images. There are however, a few obvious errors in the height model, and these errors manifest themselves in figure 8 as bright points on the side facing the radar and as horizontal dark lines (shadow) on the opposite side. The algorithms described in this paper do not resolve the ambiguity between slope and σ_0 , but only finds the most consistent set of slopes that could have resulted in the input images. As a result slopes could be adjusted due to a variation in σ_0 as opposed to a

change in the slope, resulting in false heights.

Another artifact that was observed was the ringing effect along the borders of the derived model and it is anticipated that using appropriate window functions during the processing will minimize effect.

5. CONCLUSIONS AND FUTURE PROSPECTS

The above results demonstrate the advantages of using multiple SAR images for shape from shading to supplement ordinary radargrammetry. The possibility exists for combining these methods with other multiple image methods, such as photometric stereo [Hom,80]. The latter method independently solves the simultaneous set of shading equations, for each registered pixel in a multiple image data set.

During the Magellan Mission to Venus, the SAR images are expected to be of low resolution and hence the parallax induced distortions will be minimal or nonexistent in the overlapping areas. Due to the expected lack of any significant parallax information in these images, stereo techniques will not provide an accurate height model and the overlapping images can be easily registered.

For such a scenario, shape from shading provides a hope for obtaining terrain information. Furthermore, the use of multiple images should decrease the sensitivity to noise due because of the additional information that is available.

REFERENCES

1. Frankot, R., Chellapa, R., "Application of a Shape from Shading Technique to Synthetic Aperture Radar," Proc. Int'l Geoscience and Remote Sensing Symposium, Ann Arbor, MI, May 1987.
2. Horn, B., Robot Vision, MIT Press, 1986.

3. Witte, R., "Radarelinometry for the Venus Radar Mapper," Photogrammetric Engineering and Remote Sensing, Vol. 52, #1, Jan 1986.

4. Thomas, J., et al, "Multiple SAR Image Shape From Shading", in progress.

5. Baltes, H.P.ed., Inverse Scattering Problems in Optics, Springer, 1980



Fig. 1



Fig. 2



Fig. 3



Fig. 6



Fig. 4

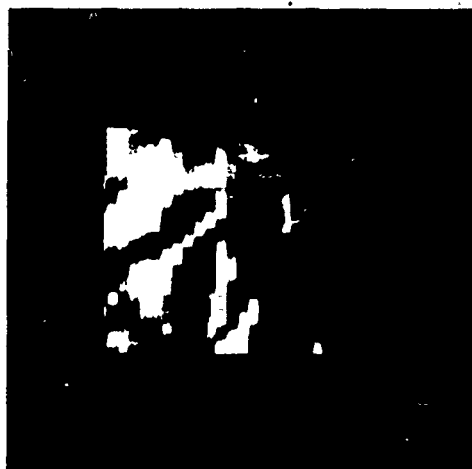


Fig. 7

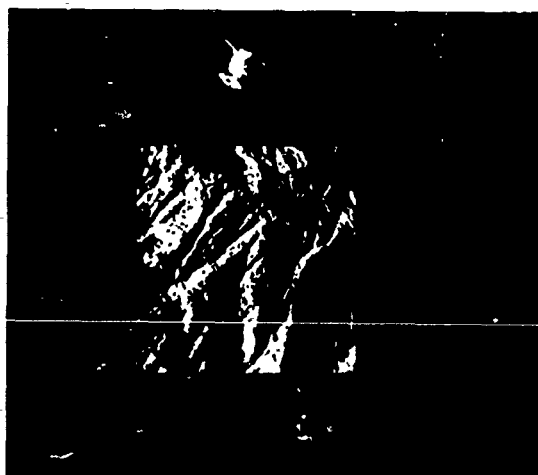


Fig. 5

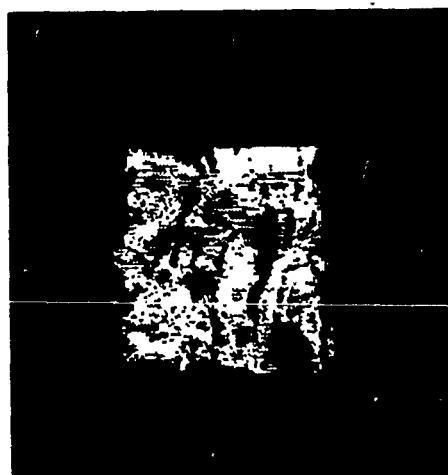


Fig. 8

DEVELOPMENT OF A SHAPE-FROM-SHADING TECHNIQUE FOR THE EXTRACTION OF TOPOGRAPHIC MODELS FROM INDIVIDUAL SPACEBORNE SAR IMAGES

B. Guindon

Canada Centre for Remote Sensing
1547 Merivale Road, Nepean, Ontario, Canada K2G 3J4

ABSTRACT

Shape-from-shading is a technique to extract the three dimensional shape of a reflecting surface based on observed image radiometric variations. In the case of remote sensing, the technique presents a potential alternative to classical stereo matching for the generation of digital elevation models and rectified imagery.

Synthetic aperture radar (SAR) imagery of rugged terrain is an attractive data source for this technique since image radiometry is dominated by terrain orientation rather than surface cover variability. This observation has been quantitatively confirmed by correlation analysis comparing real and simulated SEASAT scenes, the latter having been generated from digital elevation models assuming a universally applicable backscatter law (e.g., Guindon and Maruyama, 1986).

The present paper addresses some practical aspects related to the extraction of topographic information from single SAR scenes. A data driven approach is taken beginning with a quantitative analysis of the relationship between image grey level and terrain orientation. It is observed that image radiometry is a sensitive indicator of the range-directed component of terrain slope. Based on this fact, a new shape-from-shading technique is proposed in which radiometry is used to generate independent elevation profiles for each image range line.

Experiments have been carried out using a SEASAT scene of mountainous terrain to illustrate the above techniques. Evaluations of topographic accuracy have been accomplished through a comparison of elevation profiles extracted from the imagery and corresponding information on available 1:50,000 scale maps.

INTRODUCTION

A visual comparison of SAR imagery of rugged terrain and topographic maps suggests that image radiometry is dominated by some measure of terrain orientation. This has led a number of authors to suggest that the extraction of digital terrain models could be feasible using shape-from-shading techniques (Willey, 1986; Frankot and Chellappa, 1987).

Traditionally, shape-from-shading techniques have been based on a surface reflectance law which is assumed to be known a priori (e.g., Pentland, 1984; Horn and Brooks, 1986) such as a Lambertian model. A major

thrust of shape-from-shading research has therefore been directed at resolving the incidence angle to slope, aspect ambiguity in order to derive a consistent estimate of the three-dimensional description of the reflecting surface.

In this paper, shape-from-shading results are presented for a study of a SEASAT scene of the Adams Lake, British Columbia region. Quantitative analyses are described of comparisons between terrain slope, as derived from 1:50000 scale topographic maps, and image grey level. It is observed that image grey level for land targets is highly correlated with the range component of terrain slope. This implies that individual SAR range lines can be independently integrated to generate elevation profiles. A quantitative comparison of such shape-from-shading profiles and corresponding profiles derived from maps is made. The implications of image speckle and SAR viewing geometry on slope and profile accuracy are discussed.

DATA SET

The results presented here are based on the study of a subswath section of a SEASAT subscene, covering an area of approximately 50km x 50km near Adams Lake, British Columbia. Two example 25km x 25km sub-images are shown in Figures 1(a) and (b). The large lake labelled A in Figure 1(a) is Adams Lake (latitude: 119°30'W, longitude: 51°15'N). The topography of the region is moderate to rugged with elevations ranging from 400 to 2000 meters above sea level. The elevation contour information on available 1:50000 scale maps indicates that there exist slopes in excess of 20 degrees, implying that layover should be present in the scene.

The SEASAT scene was processed on the CCRS Generalized SAR Processor (GSAR) (Davis and Princz, 1981) in a nominal ground range-azimuth projection (i.e. without relief displacement correction) with a pixel spacing of 12.5 meters and a resolution of 25 meters. This amplitude image has 16-bit dynamic range. Further processing was applied, including radiometric compression to an 8-bit dynamic range and resampling to 100 meter resolution, 50 meter pixel spacing in order to reduce speckle.

MATHEMATICAL PRELIMINARIES

Before going on to the analysis of real SAR data, it is useful to present a simple mathematical framework for discussion. The mathematical notation of Willey,

1986 will be used here.

The coordinate system of the input SAR image will be noted by (x', y') where x' is nominal ground range and y' is azimuth. The terrain surface can be characterized by the three dimensional coordinate grid (x, y, z) , where, for convenience, y is aligned to the image azimuthal direction, x is the true ground range and z the local vertical.

Since the parallax shift in SAR imagery is in the range direction only, the input image and surface coordinate systems are related by the equations

$$\begin{aligned} y' &= y & (1) \\ x' &= x + z(x, y) \cdot e_z / e_x & (2) \end{aligned}$$

where e_x, e_z are the ground range and vertical components of a unit vector pointing from the surface back to the radar. The second term on the right side of equation 2 represents the parallax shift. Similarly, image foreshortening can be quantified by the derivative

$$\frac{\partial x'}{\partial x} = \frac{1}{1 + e_z \frac{\partial z}{\partial x}} \quad (3)$$

where $\partial z / \partial x$ is the range component of terrain slope. In non-layover conditions x' will increase with increasing x . For the case of a terrain slope equalling the view direction the denominator on the right side of equation (3) goes to zero. For steeper slopes x' will decrease with increasing x .

SLOPE INFORMATION CONTENT OF SAR RADIOMETRY

A qualitative comparison of the imagery and contour lines on 1:50000 scale maps suggests that the levels of topographic detail of the two data sources are comparable. Map features such as ridges and valleys which exhibit changes in terrain slope and/or aspect are in general detectable on the SEASAT image by the presence of localized radiometric variability. Conversely, land regions which are characterized on the map by constant slope and aspect in turn exhibit approximately constant image grey level. Given this correspondence in level of detail, 1:50000 scale maps have been used as a primary source of topographic 'truth' in correlation studies relating image radiometry and topographic descriptors.

A quantitative investigation has been undertaken through the identification and analysis of 'radiometric tie points', hereafter referred to as RTPs. A RTP is defined as a point which is identifiable both on the topographic map and the image and which lies in an extended region of constant slope and aspect (as indicated by the map contours). A total of 125 RTPs have been identified on the Adams Lake scene. From the map contours and a knowledge of the SAR imaging geometry, various incidence angle measures described below can be estimated. Statistical comparisons between these measures and RTP grey level have then been carried out.

(a) Grey Level Versus Total Incidence Angle

Traditional shape-from-shading methodologies assume grey level is indicative of the radiation incidence angle. Figure 2 shows the relationship for the 125 RTPs. Some of the RTPs lie in regions of layover and have been separated out by arbitrarily assigning them a negative value. For SEASAT, the incidence angle for a horizontal surface is approximately 21 degrees.

Two conclusions can be drawn from a study of this figure.

- (i) The grey level dynamic range encompassed by the RTPs in layover is equivalent to that of RTPs on forward facing slopes not in layover. An additional ambiguity regarding terrain slope will arise because of this overlap.
- (ii) Although a trend of increasing image grey level with decreasing terrain slope is apparent for reverse slopes (i.e. $I > 21^\circ$) no similar significant correlation exists for those RTPs on forward facing slopes.

It is concluded that SAR image grey level is not an effective indicator of total incidence angle and hence, is not an accurate measure of the local terrain surface normal direction.

(b) Grey Level Versus the Range Component of Incidence Angle

SAR imagery, acquired in a steep imaging mode such as the SEASAT case, exhibits extreme geometric foreshadowing effects even for moderate relief. In addition, geometric foreshadowing goes hand in hand with observed radiometric variability. Since foreshadowing is a function only of the range component of terrain slope, it is argued that image grey level should be a more sensitive measure of the range component of incidence angle, I_r , than total incidence angle. I_r is simply the angle between the view direction vector \hat{e} and the projection of the surface normal in the local vertical-ground range plane.

Figure 3 illustrates the grey level I_r relationship for the 125 RTPs. It can be seen that, for nonlayover RTPs, the trend of increasing grey level with decreasing I_r is now combined for the forward slopes. Although an ambiguity still exists in interpreting image grey level as either a positive or a negative value of I_r , the problem of the one to many relationship arising between total incidence angle and (slope, aspect) does not arise here because the vectors defining I_r are restricted to a planar projection surface.

If we accept the model of SAR image radiometry as a measure only of the range component of slope, two important implications arise for the potential extraction of 3D surface information.

- (i) Any given range line can be treated independently, that is, the grey level of pixels on that line can be transformed to slope estimates which in turn can be integrated to provide a unique elevation profile.
- (ii) Since there is no significant information about azimuthal slope, an additional source of "azimuthal control" data is required. This arises from the fact that each range line elevation profile is only a "relative" profile since its ground range and elevation origins are arbitrary. Azimuthal control is required to tie these adjacent integrated profiles to a common and absolute origin.

ACCURACY ANALYSIS OF THE RADIOMETRIC CALIBRATION

In this section, we consider the accuracy with which terrain slope can be extracted from image grey level and the implications of this accuracy on integrated elevation profiles derived from individual

image range lines.

Since the absolute radiometric calibration of the radar image and the precise nature of the reflectance law of the terrain surface are unknown, we have resorted to polynomial regression to derive a calibration for extracting incidence angle from grey level. Rather than attempting to define a single high order polynomial curve for all nonlayover RTPs, a composite calibration has been generated made up of separate second order polynomial fits for reverse and forward facing slopes (i.e., for angle slope ranges of 20 to 70 degrees and 0 to 20 degrees). This has been done for two reasons.

- (a) Since the derived slopes will be integrated to provide elevation profiles at the nominal ground range of 50 meters, the value of 20 degrees corresponds to the transition point between undersampling and oversampling of the input data.
- (b) Residual speckle present in the imagery will affect angle estimation accuracy. Since speckle is signal dependent its influence on dark reverse slopes and bright forward slopes will be different.

Rows 1 and 2 of Table 1 summarize the parameters of the composite calibration curve. For a selected set of grey levels covering the overall radiometric dynamic range the table presents the calibration estimates of range component of incidence angle and their one sigma uncertainties. These uncertainties are a function of the experimental uncertainty in estimating angle for individual points but also reflect the limited number of available RTPs. For example, in order to achieve a calibration accuracy of 0.1 degrees, totals of 4000 and 1000 RTPs would be required for reverse and forward slopes, respectively.

Rows 3, 4 and 5 summarize some of the implications of angle estimate errors on the integration process. These data include a ground range scale factor which is defined as the ratio of the true ground range to nominal input image pixel size. Rows 4 and 5 give the ground range and vertical errors corresponding to \pm one sigma incidence angle error. Three points should be noted

- (i) The incidence angle calibration uncertainty refers to the variation in angle estimates based on a large number of calibration curves, each defined by an independent set of randomly selected RTPs. For any given calibration curve, the angle estimation error will be systematic with grey level not random.
- (ii) Although reverse slope ground range and vertical errors are small on a per pixel basis, reverse slopes will be significantly lengthened due to the SAR viewing geometry. Given the systematic nature of the calibration error, net errors accumulated when integrating up a reverse slope can be large.
- (iii) Extreme forward slopes approaching layover (i.e., within 3 to 5 degrees of layover) exhibit very large integration errors on a per pixel basis. Such slopes are also characterized by being comprised of very few pixels because of foreshortening and high levels of speckle noise.

ANALYSIS OF ELEVATION PROFILE DERIVED FROM SAR RADIOMETRY

(a) Single Profile Analysis

In this section, a brief comparison is carried out of two integrated elevation profiles, derived from SAR radiometry, and corresponding elevation profiles manually extracted from 1:50000 scale maps (see Figures 4(a) and (b)). These profiles, sampled every 250 meters and ranging in length from 10 to 20 km, were selected to encompass a range of slope gradients. They also had to begin on a feature which is well defined both on maps and imagery and with a well-defined vertical scale origin. Prominent shoreline features on the larger lines suited the bias. A visual inspection of these figures indicates that the largest errors are introduced during the integration through steep, forward slopes.

(b) Integration Stability Analysis

The elevation profile shown in Figure 4(a) is representative of the relief characterizing the terrain between East Barrier and Adams Lakes, i.e., where the elevation rises from 625 meters on East Barrier Lake to peak in the range 1500 to 1700 meters then steadily declines to the level of Adams Lake (404 meters) with mean slopes of 10 to 15 degrees, i.e. layover is not expected to be a present.

Integrated profiles were generated starting at East Barrier Lake for 80 independent range line (i.e. every second image line encompassing the lake). The statistics of the predicted Adams Lake level were found to be the following:

mean level = 498 meters (map = 404 meters)
standard deviation = 234 meters
standard error in the mean = 27 meters

Since the difference between mean predicted and map level is almost four times the standard error in the mean, we conclude that there is a significant bias in the radiometric calibration.

To demonstrate the integration sensitivity on calibration variations, two additional calibration curves were presented in which the forward slope calibration was changed slightly from a predicted incidence angle of 9.3 degrees at grey level 120 by \pm 0.2 degrees. This results in a change in mean lake level of \pm 80 meters which indicates that the mean predicted level is consistent with the estimated accuracy of the available calibration.

The large variation in the predicted mean (237 meters) indicates that random speckle can cause significant integration errors. This suggests that additional azimuthal control will be required to periodically correct profiles. The requirements of such control are under investigation and will be published elsewhere.

CONCLUSIONS

An analysis has been undertaken of the topographic information content of SAR radiometry with a view to assessing the feasibility of shape-from-shading for terrain model extraction. For a study of a SEASAT scene of moderate relief, the following conclusions are reached.

- (i) SAR image radiometry is a strong indicator of the range component of terrain slope. It can therefore be used to derive elevation profiles for individual image range lines. Slope estimation accuracies of 1 to 2 degrees have been achieved.

- (ii) The largest integration errors occur on forward slopes where gradient approach layover conditions. Because of the steep viewing geometry of SEASAT, the accumulated integration error for regions with slopes of 10 to 15 degrees are unacceptable from a mapping point of view. Results can be expected to be improved with a more accurate calibration and/or a more sophisticated integration algorithm.
- (iii) Layover cannot be detected on the basis of grey level alone. Methods for detection based on slope trends are under investigation and will be published elsewhere.

REFERENCES

- Wilkey, R.L., 1986, 'Radarclinometry for the Venus Radar Mapper', Photogrammetric Engineering and Remote Sensing, Vol.52, pp.41-50.
- Frankot, R.T. and Chellappa, R., 1987, 'Application of a Shape From Shading Technique to Synthetic Aperture Radar Imagery', Proceedings of IGARSS'87 Symposium, pp.1323-1329.
- Guindon, B. and Maruyama, H., 1986, 'Automated Matching of Real and Simulated SAR Imagery as a Tool for Ground Control Point Acquisition', Canadian Journal of Remote Sensing, Vol.12, pp.149-158.
- Pentland, A.P., 1984, 'Local Shading Analysis', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.6, pp.170-187.
- Horn, B.K.P. and Brooks, M.J., 1986, 'The Variational Approach to Shape From Shading', Computer Graphics and Image Processing, Vol.33, pp.174-208.

TABLE 1: Composite Radiometric Calibration for the Adams Lake SEASAT Subscene (Non-layover conditions)

IMAGE GREY LEVEL

| | 50 | 70 | 90 | 120 | 150 |
|---------------------------------------|--------------|--------------|--------------|--------------|----------------|
| 1. Range Component of Incidence Angle | 72° | 42.8° | 22.3° | 9.3° | 1.5° |
| 2. Incidence Angle Uncertainty | 2.3° | 1.5° | 0.9° | 0.5° | 0.7° |
| 3. Scale Factor | 0.24 | 0.49 | 0.94 | 2.2 | 12.9 |
| 4. Ground Range Error (meters) | -0.7 +0.8 | -0.9 +1.0 | -1.8 +1.9 | +5.9 -5.3 | +559. -203. |
| 5. Vertical Error (Meters) | +1.2 -1.0 | +0.4 -0.4 | +0.7 -0.7 | +2.3 -2.0 | +215. - 78. |

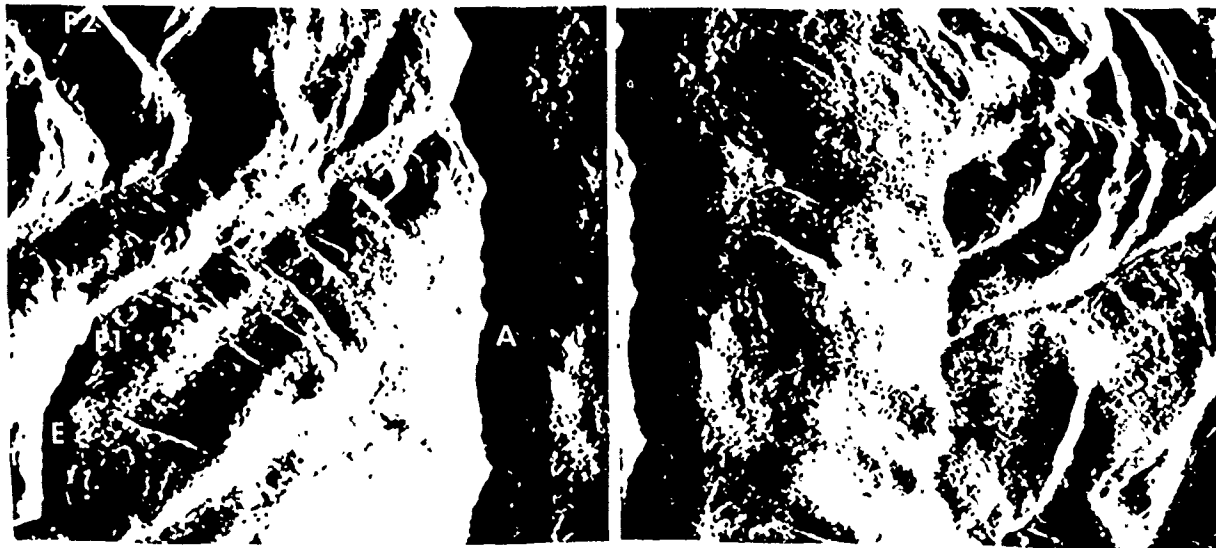


FIGURE 1: Two example subscenes of the Adams Lake region as imaged by SEASAT. The lakes labelled A and E are Adams Lake and East Barrier Lake respectively. The points P1 and P2 refer to the origins of the two example elevation profiles analyzed in the text.

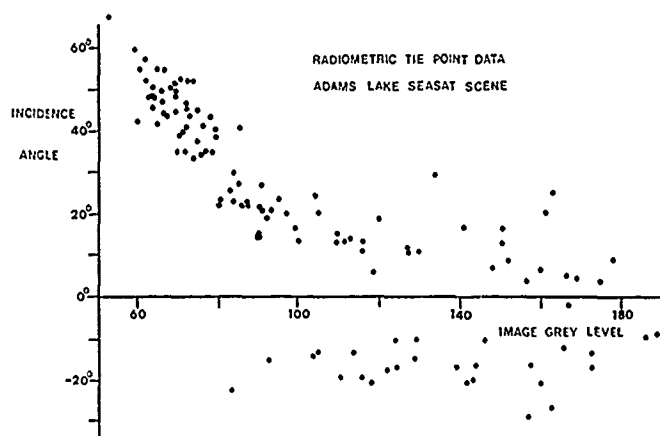


FIGURE 2: A graph of total radiation incidence angle versus image grey level for the 125 RTPs. RTPs in layover have been assigned negative values for clarity. For forward slopes image grey level and incidence angle are poorly correlated.

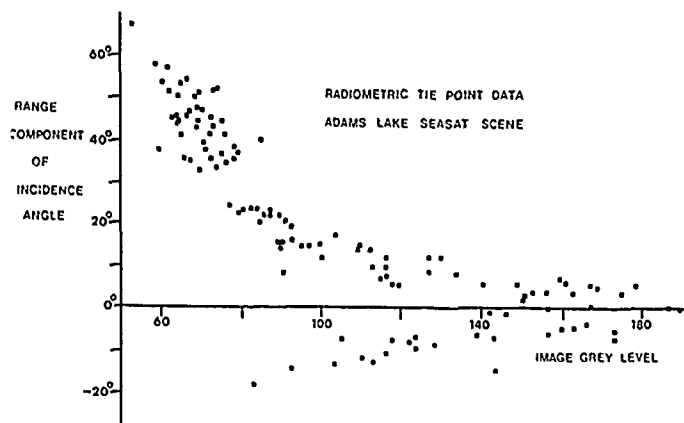


FIGURE 3: A graph of range component of incidence angle versus image grey level for the 125 RTPs. A comparison of Figures 2 and 3 indicates that SAR radiometry is a more accurate indicator of the range component rather than the total incidence angle.

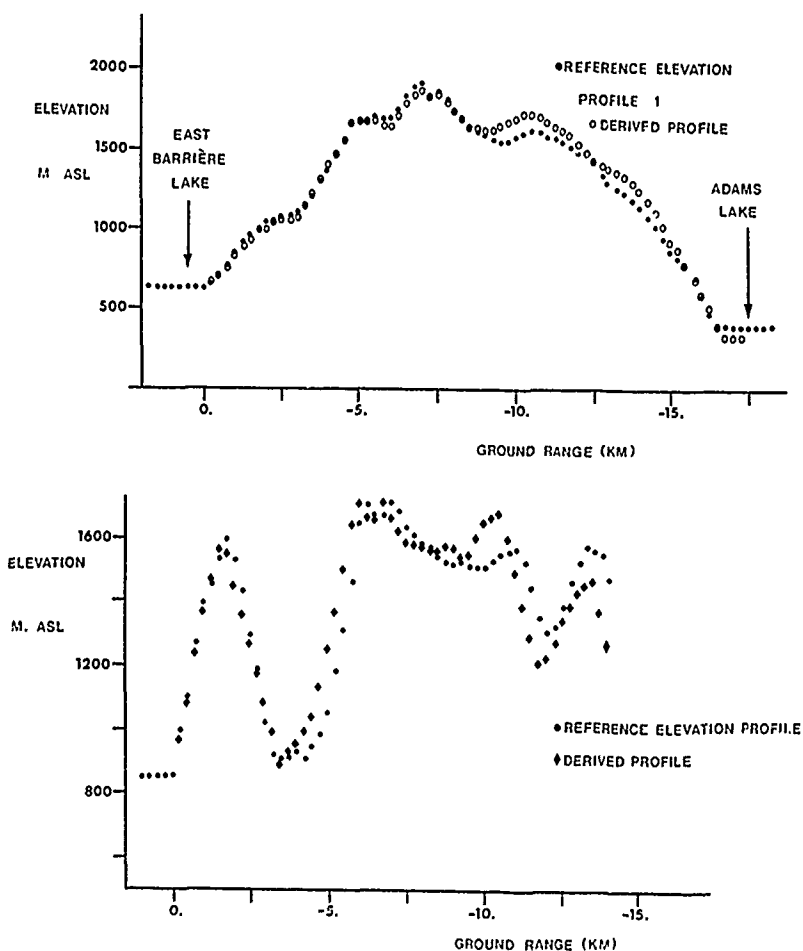


FIGURE 4: A comparison of two elevation profiles, extracted from 1:50000 maps and corresponding predicted profiles based on shape-from-shading integration.

A METHOD OF DETERMINING THE PIXEL GEOLOCATION OF SAR IMAGES IN OCEANIC AND POLAR REGIONS

U. Tretter

German Aerospace Research Establishment (DFVLR)

WT-DA/MV SAR

Muenchnerstrasse 20,

D-8031 Oberpfaffenhofen, West Germany

Telephone (0049) 8153/28-1207

Telex 5270286 DAAL D

Telefax (0049) 8153/281137

Abstract

For the ERS-1 spaceborne surface-imaging Synthetic Aperture Radar, one of the most important subtasks, due to precision image product generation, consists of solving the problem of geolocating the SAR image elements, in terms of geodetic coordinates, as precisely as possible, especially in oceanic and polar regions.

Due to the probable unavailability of ground control points (GCPs) in these areas, a new algorithmic procedure was developed, based upon the use of Range-Doppler methodology in combination with geodetic calculations, without requiring GCPs. The procedure functions independently of spacecraft attitude values and requires the following input parameters: satellite ephemeris data, the derived slant range distance, earth model describing parameter and the Doppler Centroid. As an additional option, the use of terrain heights is also foreseen. Relying on a vector model, an equation with two unknown variables is solved by ascertaining one of these in an iterative way. After several computational phases, this method calculates the geodetic position or geolocation of the image element in relation to the chosen earth ellipsoid. As an independent control of the accuracy of the algorithm, the Doppler Centroid value and the calculated Doppler value of the pixel can be compared. Any deviation reflects the degree of precision in the algorithmic procedure itself. Misgeolocation due to errors concerning the satellite ephemeris, Doppler Centroid, slant range distance and terrain height is taken into consideration in a final accuracy estimation.

To conclude, one can say that tests on actual data have shown that the achievable accuracy of the algorithm itself is convincing and will be a useful tool for geolocating SAR images. However, in a qualified sense, it has to be emphasised that the real accuracy will be a function of the inherent accuracy of the input parameter.

Keywords: SAR Processing, SAR Image Geolocation, ERS-1

A. Roth, H. Craubner, Th. Bayer

DLR, WT-DA, German Aerospace Research Establishment
D-8031 Oberpfaffenhofen, FRG

Abstract

This paper gives an overview on the existing prototype versions for the generation of geocoded products in the German Processing and Archiving Facility (D-PAF) of the ERS-1 and SIR-C / X-SAR missions. While the terrain correction is performed by an exact pixel to pixel transformation, the ellipsoid corrected products are generated using block processing. As a first step this requires an algorithm which provides a spatial transformation model from a rectangular grid, defined on an output image, to an irregular grid, defined on the input SAR slant range image. It was assumed that image specific approximating polynomials for the orbit and the SAR Doppler shift, depending on the pixel location (range and azimuth) are available.

The block processing is performed by an efficient pixel by pixel transformation algorithm including cubic convolution as resampling method. This procedure resamples the input SAR slant range image to a certain map projection in three steps using two bilinear transformations and two image shears. This implies an one dimensional resampling method for cubic convolution.

Keywords:

Ellipsoid Correction, Terrain Correction, Without Visual Inspection, One-dimensional Block Processing, Cubic Convolution, Doppler Reference Function, German PAFs for ERS-1 and SIR-C/X-SAR Missions.

1. Introduction

Traditionally, geometric distortions in remote sensing data have been corrected by manually located tiepoints, calculation of a transformation polynomial and a resampling algorithm (Haberäcker, 1985). This paper describes a geocoding system linked to the output of a SAR preprocessing system, which has the capability of producing geocoded images without any visual inspection of the raw image (like searching controlpoints). In case of ellipsoid correction gridpoints are calculated in input as well as output images based on the exact solution of two equations, the SAR Doppler equation and the SAR range equation (Kwok, 1987). The pixels located between the grid lines will subsequently be resampled by bilinear transformation, using the four neighbouring gridpoints as reference locations. This part has originally been developed for scanning systems by Friedmann (Friedmann, 1981). Therefore these formulas can also be used for the rectification of optical data.

The root idea is the replacement of the higher order polynomial transformation acting in two dimensions by an one-dimensional procedure. The results of this transformation, rectification in lines and columns as well as a rotation, are performed within three steps separating the rectifications and replacing the rotation by two image shears (Curlander, 1987). This procedure allows the parallelizing of calculations for the utilization of an array processor. The algorithmic computational efficiency mainly consists of the one-dimensional cubic convolution interpolation.

2. Geometric Model

There are two basic concepts for the geocoding of SAR images. The terrain correction considers the earth's relief described by a Digital Elevation Model (DEM), while ellipsoid correction takes into account the smooth surface of a defined ellipsoid. So the basic geometric model is identical for both procedures. The only difference is that for the calculation of the earth's surface point vector \vec{p} in the case of terrain correction the height above the geoid is considered. This implies a pixel by pixel procedure, while the generation of ellipsoid corrected products only requires the calculation of a transformation grid.

2.1 Conversion from Map Coordinates to Cartesian

System

As a first step the map coordinates defining the location of a point in the output image have to be transformed to a Cartesian coordinate system, which originates at the center of the geoid and follows the earth rotation. The appropriate transformations are well described elsewhere and are not discussed in this paper.

There will be a choice of various projections (see also Graf, 1988):

1. Transverse Mercator Projection
2. Oblique Mercator Projection
3. Lambert Conformal Projection
4. Stereographic Projection (oblique and polar aspect)
5. Mercator Projection

The result of the mapping algorithm is a three-dimensional net covering the earth's surface. Each netpoint is defined in its three coordinates: x, y, z . In the case of terrain correction this net consists of all image points.

2.2 SAR Pixel Allocation

Starting from a defined location on the earth's surface, given in Cartesian coordinates the accurate location of the corresponding pixel on the SAR image has to be determined. (Negative column or line numbers are allowed. In this case the earth location is not mapped inside the specific image, but to a virtual place outside.)

Two different equations have to be solved simultaneously to find the column and line of the SAR pixel. One equation approximates the SAR Doppler shift of a point target and the other one describes the dependency of the slant range between sensor and earth location from the line number of this pixel.

A set of image specific constants, which have to be computed during the preprocessing, are input to the algorithm.

The SAR image coordinates, which should be calculated, are called (i, j) , where i is the column number or azimuth coordinate and j is the range

coordinate or line number. Both numbers need not be an integer. (The mapped earth location $[x, y, z]$ may correspond to a place between two lines and two columns of the SAR image.)

x, y and z are the coordinates of one special netpoint in a geocentric Cartesian coordinate system, which rotates with the earth. (x, y, z) may be expressed in vector notation:

$$(1) \quad \vec{p} = [x, y, z]$$

The velocity of the point $[x, y, z]$ relative to an inertial system may be calculated as:

$$(2) \quad \dot{\vec{p}} = \vec{\omega}_E \times \vec{p}$$

where $\vec{\omega}_E$ is the earth rotational velocity:

$$(3) \quad \vec{\omega}_E = \begin{bmatrix} 0 \\ 0 \\ 2\pi/24 \text{ h} \end{bmatrix}$$

2.2.1. SAR Doppler Equation. The equation describes the Doppler frequency shift of a point target as seen from the satellite (Li, 1983):

$$(4) \quad F_1(i, j) = f_{DC} - \frac{2 \cdot (\vec{p} - \vec{s}) \cdot (\vec{p} - \vec{s})}{\lambda \cdot |\vec{p} - \vec{s}|} = 0$$

$$\text{with: } f_{DC} = f_{DC}(i, j)$$

$$\text{and: } \vec{s} = \vec{s}(i), \quad \dot{\vec{s}} = \dot{\vec{s}}(i)$$

where f_{DC} is the Doppler frequency shift and λ is the radar wave length. \vec{s} is the location of the sensor during the exposure of the specific image pixel and $\dot{\vec{s}}$ is the velocity of the sensor along the orbit, defined in the same Cartesian, geocentric coordinate system as \vec{p} . λ is a fixed parameter, while f_{DC} changes along azimuth and range direction. The orbit location as well as the velocity are time dependent vector functions. Because azimuth is a linear function of time, both vectors describing the orbit may be considered as functions depending on pixel azimuth.

In our case it was assumed, that f_{DC} and \vec{s} can be approximated by polynomials:

$$(5) \quad f_{DC} = d_0 + d_1 \cdot j + d_2 \cdot j^2 + d_3 \cdot i + d_4 \cdot i^2$$

$$(6) \quad \begin{cases} x_i = a_0 + a_1 \cdot i + a_2 \cdot i^2 + a_3 \cdot i^3 \\ y_i = b_0 + b_1 \cdot i + b_2 \cdot i^2 + b_3 \cdot i^3 \\ z_i = c_0 + c_1 \cdot i + c_2 \cdot i^2 + c_3 \cdot i^3 \end{cases}$$

The polynomial coefficients $a_0, a_1, \dots, b_0, b_1, \dots, c_0, c_1, \dots$ and d_0, d_1, \dots are image specific and have to be calculated by a preprocessing unit before the algorithm is started. The accuracy for the orbit polynomial is known to be good enough for that small part of the orbit considered as the synthetic aperture.

2.2.2. SAR Range Equation.

$$(7) \quad F_2(i, j) = r_0 - m_r \cdot j - |\vec{p} - \vec{s}|$$

The equation describes the linear dependency of the slant range between sensor and earth location on the line number of a SAR pixel (= range coordinate), where r_0 and m_r are scaling constants. (r_0 is image specific, while m_r is sensor specific.) Because \vec{s} depends on azimuth i , F_2 is a function of both variables i and j .

Both equations F_1 and F_2 are nonlinear in the unknown variables i, j . Therefore, it is impossible to solve the equation system directly.

2.3. Iteration Method

Because there is no direct solution to the problem, a start value (i_0, j_0) corresponding to a netpoint $[x, y, z]$ has to be found. We assume that the geographic coordinates of the SAR image center are available in the annotation data produced by the preprocessing unit. This location serves as a first tiepoint. After improving (i_0, j_0) to a specified accuracy an iteration process applying the result may serve as starting value for the next gridpoint in the neighbourhood of the previous one. This procedure has to be repeated until all gridpoints are calculated.

The Newton iteration method with partial derivatives was applied:

$$(8) \quad \begin{cases} F_1(i_0, j_0) + \frac{\partial F_1}{\partial i} \cdot \Delta i + \frac{\partial F_1}{\partial j} \cdot \Delta j = 0 \\ F_2(i_0, j_0) + \frac{\partial F_2}{\partial i} \cdot \Delta i + \frac{\partial F_2}{\partial j} \cdot \Delta j = 0 \end{cases}$$

where:

$$(9) \quad \begin{aligned} \frac{\partial F_1}{\partial i} &= \frac{\partial f_{DC}}{\partial i} - \frac{\partial}{\partial i} \left[\frac{2 \cdot (\vec{p} - \vec{s}) \cdot (\vec{p} - \vec{s})}{\lambda \cdot |\vec{p} - \vec{s}|} \right] \\ &= d_3 + 2 \cdot d_4 \cdot i^2 \\ &+ \frac{2}{\lambda} \cdot \left[\frac{\frac{\partial \vec{s}}{\partial i} \cdot (\vec{p} - \vec{s}) - (\vec{p} - \vec{s}) \cdot \frac{\partial \vec{s}}{\partial i}}{|\vec{p} - \vec{s}|} \right] \\ &+ \frac{((\vec{p} - \vec{s}) \cdot (\vec{p} - \vec{s})) \cdot (\frac{\partial \vec{s}}{\partial i} \cdot (\vec{p} - \vec{s})) - ((\vec{p} - \vec{s}) \cdot \frac{\partial \vec{s}}{\partial i}) \cdot ((\vec{p} - \vec{s}) \cdot (\vec{p} - \vec{s}))}{(\vec{p} - \vec{s})^2 \cdot |\vec{p} - \vec{s}|} \end{aligned}$$

$$\text{with: } m_r = \frac{dj}{dt}$$

(m_r : scale factor,
linear dependence of azimuth on time)

$$(10) \quad \left\{ \frac{\partial F_1}{\partial j} = \frac{\partial f_{DC}}{\partial j} = d_1 + 2 \cdot d_2 \cdot j \right.$$

$$(11) \quad \left\{ \frac{\partial F_2}{\partial i} = - \frac{(\vec{p} - \vec{s}) \cdot \frac{\partial \vec{s}}{\partial i}}{|\vec{p} - \vec{s}|} \right.$$

$$(12) \quad \left\{ \frac{\partial F_2}{\partial j} = m_r \right.$$

and:

$$(13) \quad \left\{ \frac{\partial \vec{s}}{\partial i} = \begin{bmatrix} a_1 + 2 \cdot a_2 \cdot i + 3 \cdot a_3 \cdot i^2 \\ b_1 + 2 \cdot b_2 \cdot i + 3 \cdot b_3 \cdot i^2 \\ c_1 + 2 \cdot c_2 \cdot i + 3 \cdot c_3 \cdot i^2 \end{bmatrix} \right.$$

$$(14) \quad \left\{ \frac{\partial^2 \vec{s}}{\partial i^2} = \begin{bmatrix} 2 \cdot a_2 + 6 \cdot a_3 \cdot i \\ 2 \cdot b_2 + 6 \cdot b_3 \cdot i \\ 2 \cdot c_2 + 6 \cdot c_3 \cdot i \end{bmatrix} \right.$$

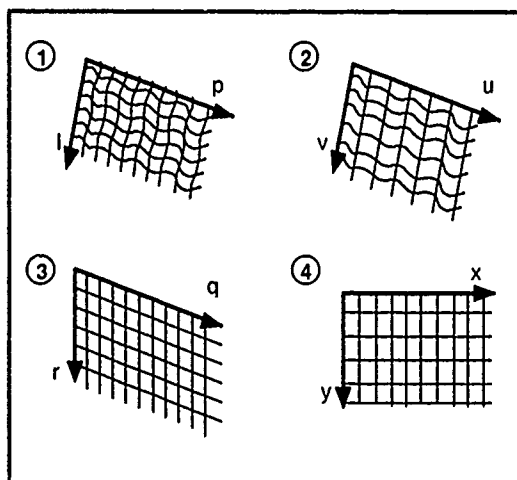
Inserting (i_0, j_0) into equations (9) to (10), we get a linear equation system (8) for the unknowns $(\Delta i, \Delta j)$, which we can solve directly. After adding $(\Delta i, \Delta j)$ to (i_0, j_0) , the procedure may be repeated until the accuracy requirements are fulfilled.

3. Generation of Ellipsoid Corrected Geocoded Product

3.1 Algorithmic Characteristics

The parameters for a higher order polynomial transformation contain rectification in lines and columns and a rotation. For this algorithm these parts are separated and will be performed in several steps.

The results of the different steps are demonstrated in figure 3.1, where 1 shows the input, 2 and 3 show the intermediate images and 4 shows the final output. The first step 1 to 2 performs the rectification line by line and additionally contains the oversampling. Then (2 to 3) rectification in columns and first image shear are performed. Finally (3 to 4), the second image shear completes the geocoding.



- figure 3.1 -

The following chapters deal with the replacement of the direct transformation by two one-dimensional rectifications and image shears.

3.2 Rectification

The rectification will be performed by transforming the grid and using two bilinear interpolations within each subimage. This implies an extraction of the one-dimensional transformation parameters e, f out of parameters a and b , which characterize the bilinear polynomial. Parameters e and f can be calculated by solving equations (15) and (16) and neglecting the quadratic terms.

The equations for the two-dimensional bilinear interpolation are:

$$(15) \quad l = a_0 + a_1 x + a_2 y + a_3 x y$$

$$(16) \quad p = b_0 + b_1 x + b_2 y + b_3 x y$$

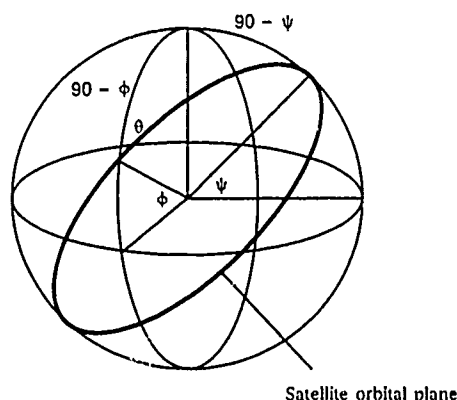
So the final equations are:

$$(17) \quad l = f_0 + f_1 x + f_2 y + f_3 x y$$

$$(18) \quad p = e_0 + e_1 x + e_2 y + e_3 x y$$

3.3 Rotation

The rotation angle for the orientation of a scene to north varies, depending on the latitude, due to the inclination of the orbital plane. From figure 3.2, a formula can be derived to estimate this angle.



- figure 3.2. -

$$(19) \quad \theta = \arcsin \frac{\cos \psi}{\cos \phi}$$

ψ = inclination

ϕ = latitude of subimage center

Equations (20) and (21) describe the separation of a two-dimensional left rotation into two one-dimensional image shears. Numbers (22.1) and (22.2) show the scalar description of these rotations which representing the formulas for the two steps.

Left Rotation

$$(20) \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$$

$$(21) \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \tan \theta & \sec \theta \end{pmatrix} \cdot \begin{pmatrix} \cos \theta & -\sin \theta \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$$

$$(22.1) \quad \begin{aligned} x' &= q \\ y' &= q \tan \theta + r \sec \theta \end{aligned}$$

$$(22.2) \quad \begin{aligned} q &= x \cos \theta - y \sin \theta \\ r &= y \end{aligned}$$

θ = rotation angle

3.3 Three Step Algorithm

The results of chapters 3.1. and 3.2. are combined to a three step procedure. To avoid aliasing effects, oversampling is performed. The oversampling factor is calculated by

$$(23) \quad n_f = 1 + \tan \theta$$

Equations (24) to (26) describe the three steps, while figure 3.1 shows the corresponding effects of the different calculations.

$$(24.1) \quad l = v$$

$$(24.2) \quad p = \theta_0 + \theta_2 v + \frac{(\theta_1 u + \theta_3 u v)}{n_f}$$

$$(25.1) \quad y' = \frac{q}{n_f} \tan \theta + \frac{r}{\cos \theta}$$

$$(25.2) \quad v = f_0 + f_1 \frac{q}{n_f} + f_2 y' + f_3 \frac{q}{n_f} y'$$

$$(25.3) \quad u = q$$

$$(26.1) \quad q = n_f (x \cos \theta - y \sin \theta)$$

$$(26.2) \quad r = y$$

In the first step (24), rectification in direction of ascending columns is done. The second step combines the second rectification in line direction and the first image shear (25), while the last step (26) performs the second image shear.

3.4 Cubic Convolution Interpolation

In general resampling means the assignment of a grey value to a discrete given pixel. It can be seen in equations (24) to (26) that indirect resampling is used, i.e. the one dimensional calculation of the pixel position in the input image vector out of the discrete given pixel location in the output image vector. Therefore, the gray value $O(t)$ must be calculated from an amount of input pixels $I(1), \dots, I(n)$. I and O are given as vectors where vector I has the sampling period T and O T' . The sampling period T can be understood as a given distance between pixels within the input image. T' is the desired spacing for the output image. $O(t)$ is the discrete given pixel location which leads to a pixel position in vector I located between $I(n)$ and $I(n+1)$, i.e. $nT \leq mT' \leq (n+1)T$.

$O(t)$ is calculated by

$$(27) \quad O(t) = \sum_{i=n-1}^{n+2} I(i) \cdot c \left(\frac{mT' - nT}{T} \right)$$

$c(t)$ is the cubic convolution kernel which is defined as

$$(28) \quad c(t) = \begin{cases} \frac{3}{2} |t|^3 - \frac{5}{2} |t|^2 + 1 & 0 \leq |t| \leq 1 \\ -\frac{1}{2} |t|^3 + \frac{5}{2} |t|^2 - 4|t| + 2 & 1 \leq |t| \leq 2 \\ 0 & 2 \leq |t| \end{cases}$$

(Keys, 1981 and Pich, 1986)

The kernel for this four point interpolation is hardcoded in a lookup table, separated in 16 parts. For the calculation of the actual grey value this method takes considers the current, the previous and the two following pixels.

4. References

1. Curlander J.C., "A post-processing system for automated rectification and registration of spaceborne SAR imagery", Int. J. Remote Sensing, Vol. 8, No. 4, p. 621., 1987.
2. Friedmann D.E., "Operational resampling for correcting images to a geocoded format", 15th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, p. 195., 1981.
3. Graf Ch., Nüesch D., Meier E., Frei U., "Map Projections for SAR - Geocoding", Technical Note (ERS-D-TN-22910-A/9/88), Zuerich, September 1988.
4. Haberäcker P., "Digitale Bildverarbeitung", chapter 10.4, Carl Hanser Verlag, München Wien, 1985.
5. Keys R.G., "Cubic Convolution Interpolation for Digital Image Processing", I.E.E.E. Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-29, No. 6, p. 1153., 1981.
6. Kwok R. and Curlander J.C., "Automated Rectification and Geocoding of SAR Imagery", Proceedings of IGARSS'87 Symposium, Ann Arbor, 18-21 May 1987.
7. Li F.K. and Johnson W.T.K., "Ambiguities in Spaceborne Synthetic Aperture Radar Systems", IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-19, No. 3, May 1983.
8. Pich M., "Interpolation Task, Design Manual", DFVLR Internal Description, 1986.

AUTOMATED ICE MOTION TRACKING AT THE ALASKA SAR FACILITY

Ronald Kwok
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Abstract

A Geophysical Processor System (GPS) is being developed for routine extraction of geophysical data from SAR images acquired and processed at the Alaska SAR Facility (ASF) which will be located in Fairbanks, Alaska. To support studies in ice dynamics, one of the functions of the GPS will be to derive ice motion data from time-sequential imagery. A totally automated system is being developed for routine generation of such motion data. To accomplish the determination of ice motion, the GPS will utilize a complex ice motion estimation, segmentation and tracking algorithm. The ice motion estimator utilizes wind and buoy data to estimate the mean motion of ice fields within an image and guides the database search for location of the corresponding image containing approximately the same ice field at a different time. In order to track the complex motion of ice floes (rotation and translation) in the central basin and the ice margin, a hybrid feature tracking and area tracking approach is taken. The feature tracker computes motion of segmented region boundaries and the estimated motion from this stage is used to initialize an area correlation algorithm which generates motion data on a uniformly-spaced geocoded grid. The motion data describe the translation and rotation of the area in the neighborhood of the grid nodes. The system is implemented on a high-performance workstation equipped with an array processor. The output data products are available for distribution through the Archive and Operations System of the ASF.

Keywords - Ice Motion Tracking, Feature Extraction, Feature Matching, Area Matching, Automated System.

This work was carried out under contract with the National Aeronautics and Space Administration at the Jet Propulsion Laboratory, California Institute of Technology.

I. Introduction

The Alaska SAR Facility (ASF), to be located in Fairbanks, Alaska, will receive, record, process and disseminate SAR data acquired by the European ERS-1, Japanese ERS-1 and Canadian RADARSAT satellites. The Alaska SAR Facility (ASF) will be the first fully integrated system for handling large volumes of SAR data ever built by NASA. It basically consists of three subsystems: 1) the receiving ground station; 2) the high-speed SAR processor; and 3) an archive and operations system that supports a geophysical processor and an image analysis workstation. The science objectives of the ASF program are multidisciplinary including studies in air-sea-ice interaction, oceanography, glaciology, geology, hydrology and ecology. To support these studies, a Geophysical Processor System (GPS) is being developed to enable manipulation and analysis of the ASF data as well as routine generation of geophysical products. Presently the GPS has two components, an Ice/Oceans Processor and an image analysis workstation. This paper focuses on the description of one of the functions of the Ice Processor GPS, the generation of ice motion products from time-sequential SAR images. The products include sea ice kinematics data sampled on a pre-defined spatial and temporal grid with embedded meteorological data acquired from national weather centers.

Recently, a number of automatic ice motion tracking algorithms [Fily and Rothrock, 1986; Collins and Emery, 1988; Vesecky et al, 1988] have appeared in the literature. Most are based on the use of area correlation as an image matching technique for deriving the translation of ice floes. Area correlation is mildly tolerant to moderate rotation of image features but give erroneous results when large rotations are present. Lee [1988] devised an

operator-assisted technique to initialize the area-based search for image correspondence. Vesecky et al [1988] proposed a edge-based matching technique for constructing the translation and rotational parameters however only a sparse motion field was generated. Here, a hybrid algorithm which includes a feature matching and area matching stage is described. The algorithm seems to be fairly robust when tested with SEASAT motion pairs. The system which will be installed at ASF will be totally automated for routine generation of ice motion products.

2. Algorithm Overview

The algorithm flow is shown in Fig. 1. A description of each of the stages follows:

Ice Motion Estimator. The primary function of this estimator is for reducing the search space of the location for the image pairs containing approximately the same ice features. The estimator uses motion measurements from buoy data and wind data.

Feature Extraction. Region analysis is applied to extract the boundaries separating regions of similar intensity and texture. Region centroids are determined by an unsupervised clustering scheme and the pixels are separated using a minimum distance classifier. Fig. 2 shows a SEASAT image pair and the regions extracted are shown in Fig. 3.

Feature Matching. The region boundaries are then vectorized and transformed into a shape representation where rotations are mapped into translations in the ordinate axis as shown in Fig. 4. This shape descriptor is known as a *psi-s* curve. *Psi* is the angle made between a reference line and the tangent to the region boundary and *s* is the distance along the boundary. These curves are matched using a simple 1-d cross-correlation of short segments of the shape descriptor extracted from the two images. The highlighted region boundaries extracted from the image pair are shown in Fig. 4a and the corresponding matched *psi-s* curves are shown in Fig. 4b. Figure 5 shows the matched region boundaries generated using this feature matching approach. The vectors connect the matched region boundaries. Both translation and rotation are obtained with this technique.

Area Matching. The results from the preceding stage is used to initialize the area matching scheme which is required to generate the motion field sampled on a uniformly-spaced grid. A refinement of the translation as well as rotation of the image features in the neighborhood of the grid points are performed at this stage. Figs. 6a,b show the vector field and the rotation of the image patches generated at the grid points.

Consistency Checks and Filtering of Bad Matches. As an operational system, the ice motion processor must meet minimum performance criteria in terms of accuracy and consistency of the output motion fields. The algorithm includes a number of filtering processes and verification checks based on various match metric as well as motion constraints to filter out matches which are considered false.

3. System Description

System Architecture and Performance The system is based on a SUN-4 workstation with an array processor capable of 15Mflops. The expected throughput of the system is 10 motion products/day.

Data Products. The motion field will be sampled on an Earth-fixed SSM/I grid (grid point locations defined by a Polar Stereographic projection with the secant plane at 70 deg. N and the map vertical at 135 deg. E) with 5 km grid spacing. With this spacing, the number of grid points within a 100km image frame is approximately 400.

4. Summary Remarks

An automated system for construction of ice motion data from multitemporal SAR images is briefly described here. The tracking algorithm has been tested with SEASAT images and a limited amount of C-band aircraft data. Results generated using this dataset indicate fairly robust performance even near the ice margin where considerable deformation and large rotations of the image features can be observed.

5. Reference

- Collins, M. and W. J. Emery, A computational method for estimating sea ice motion in sequential Seasat synthetic aperture radar imagery by matched filtering, J. Geophys. Res., 93, 9241-9251, 1988.
- Fily, M., and D. Rothrock, Extracting sea ice data from satellite SAR imagery, IEEE Trans. Geosci. Remote. Sens., GE-24, 849-854, 1986.
- Lee, M., Sea-ice analysis software - ICEMAN, Proceeding of IGARSS, 1713-1716, 1988.
- Vesecky, J. F., R. Samadani, M. P. Smith, J. M. Daida, R. N. Bracewell, Observation of sea-ice dynamics using synthetic aperture radar images: automated analysis, IEEE Trans. Geosci. Remote. Sens., GE-26, 38-48, 1988.

Figure 1. Ice Motion Tracking Algorithm Flow

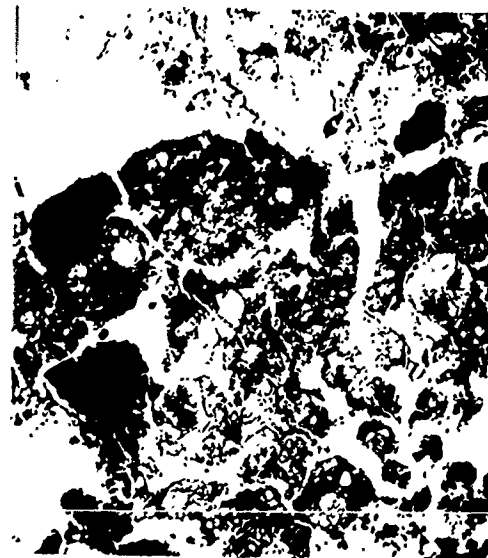
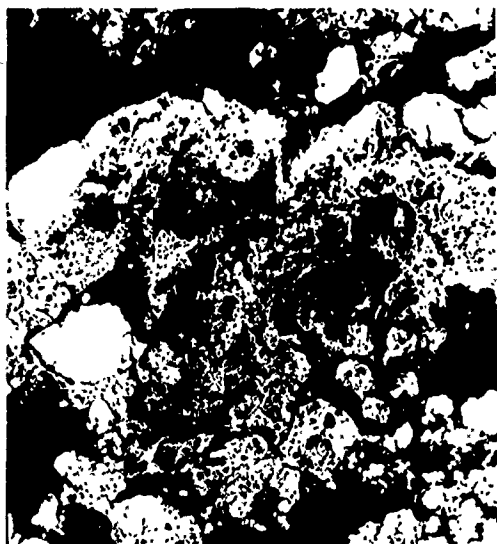
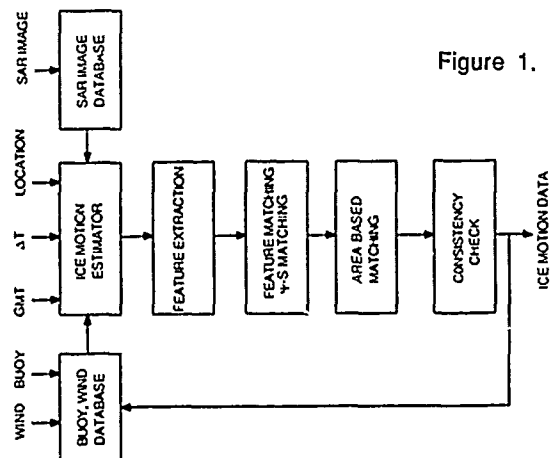


Figure 2. Original SEASAT Ice Motion Pair.

Figure 3. Binary Regions after Cluster Analysis.

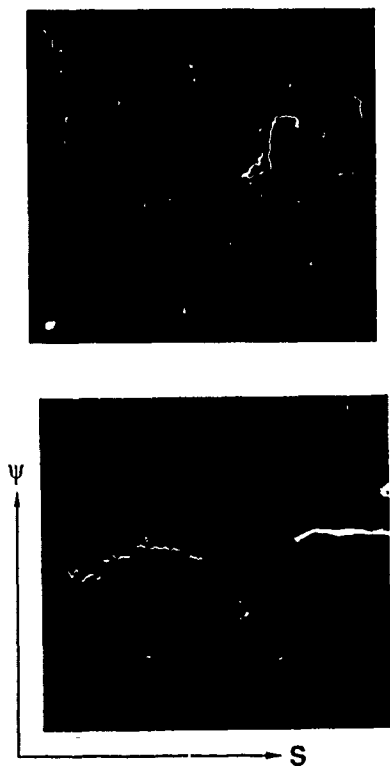


Figure 4. (a) Region Boundaries on the Image.
(b) Matched Psi-s curves.



Figure 5. Vectors showing motion of region boundaries.

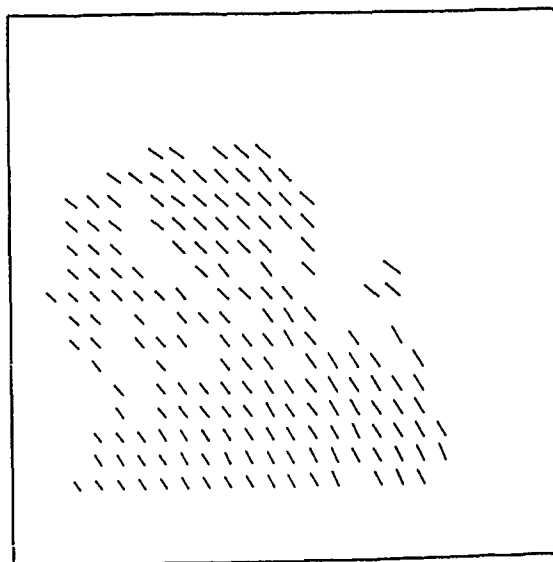
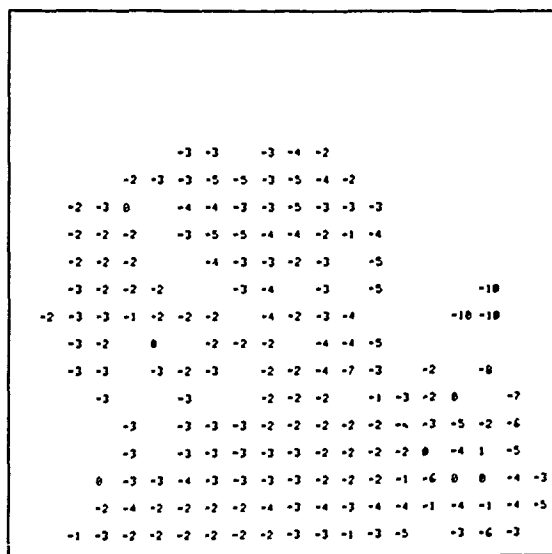


Figure 6. Output of Area Correlation. (a) Rotation angles image neighbor (b) Motion sampled on a uniformly-spaced grid.

Interpolation and sampling in SAR images

S Quegan

Department of Applied and Computational Mathematics,
University of Sheffield,
Sheffield S10 2TN, UK.

Abstract

Investigation of statistical peculiarities in Agri SAR data revealed a fundamental flaw in the resampling scheme used to generate the amplitude images. This led on to a study of how resampling affects SAR image statistics. It is shown that resampling in the complex image before forming the intensity image conserves the exponential form of the intensity image and preserves the intensity mean if the interpolation scheme has unit energy when regarded as a linear filter (eq (7)). The correlation structure of the intensity data cannot be conserved by any finite length interpolation scheme, but the correlation structure imposed by the interpolation can be explicitly calculated. Resampling in the intensity domain fails to conserve the distribution or correlation structure.

Keywords: AgriSAR, texture, speckle.

1. Introduction

The motivation for the work presented in this paper is the observation that the statistics of AgriSAR images show peculiar properties (Quegan and Yanasse, 1989). In Figure 1 we show the result of averaging pixels at constant range from a section of an AgriSAR amplitude image. The periodicity between adjacent range gates is very clear. Upon investigation this behaviour was traced to the interpolation scheme used. In order to produce approximately square pixels, the system designers of the VARAN-S system inserted an extra pixel between every directly measured value in the range direction. The first method used to do this was nearest neighbour interpolation, but this led to 'blockiness' in the data which was thought unacceptable (Freeman and Keen, 1986). As a result, the interpolation scheme was altered, as follows.

In the range direction, the VARAN-S collects 3000 complex samples n_k per azimuth position;

$$n_k = x_k + iy_k \quad (1)$$

where x_k and y_k are the in-phase and quadrature components of the signal respectively. The amplitude image is formed by taking the modulus of n_k . Between each such value formed, an extra pixel value p_k is

inserted by averaging in the complex image and taking the modulus, i.e.,

$$p_k = \left| \frac{n_k + n_{k+1}}{2} \right| \quad (2)$$

and p_k appears in the image between $|n_k|$ and $|n_{k+1}|$. Unfortunately, this scheme ignores a fundamental property of SAR data.

Over extended targets, the coherent imaging process causes x_k and y_k to be zero-mean uncorrelated Gaussian random variables (rvs) with the same standard deviation σ (we ignore texture for the moment). The intensity image formed by the operation

$$I_k = x_k^2 + y_k^2 \quad (3)$$

then has mean $2\sigma^2$, i.e. the mean of the intensity (or amplitude) image is determined by the standard deviation in the complex image. The operation (2) preserves the mean value of the complex data, but reduces its standard deviation by a factor $\sqrt{2}$, leading to the distortion of the mean values of the interpolated pixels in the amplitude data observed in Figure 1. (A more comprehensive treatment taking account of pixel correlation is given in Quegan and Yanasse (1989)).

The question arises as to what interpolation schemes are valid for SAR data, and how the type of error present in the AgriSAR data can be avoided. SAR offers a particularly interesting data-type for this problem, since we have the option of interpolating in the complex or intensity domain, and we need to understand if they are equivalent. SAR data also forces us to question what quantity we wish to preserve when we carry out resampling.

Classically, we attempt to preserve the signal itself; the Shannon-Whittaker sampling theorem then tells us to use $\frac{\sin x}{x}$ interpolation.

In practice, truncation of the interpolating sequence is necessary. The errors due to this can be quantified in both a mean-square and absolute value sense (Papoulis, 1977), but there are still unfortunate artefacts such as generation of negative values for positive signals, and distortion of

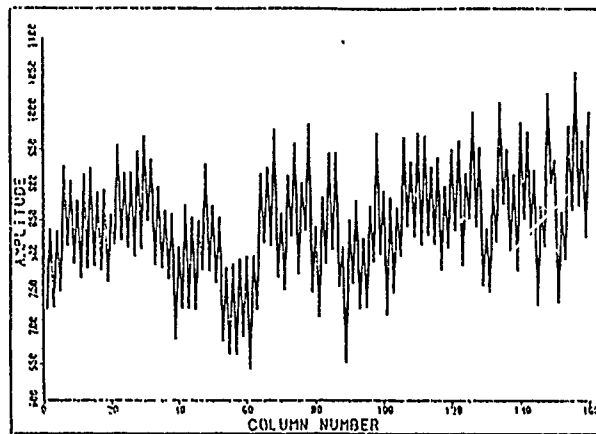


Figure 1 - Amplitude column averages for a sugar beet field

statistics (Clark et al, 1989). Some of these errors can be reduced by applying a taper to the $\frac{\sin x}{x}$ interpolator, but that is not discussed here.

Despite these problems, $\frac{\sin x}{x}$ interpolation is clearly the correct approach to use when examining SAR images in the neighbourhood of isolated scatterers, such as is necessary when examining the response from corner reflectors for calibration purposes. However, for many purposes, the most important properties of the SAR data to preserve are the statistical properties of extended targets. This approach is the one we investigate in this paper, and we set out the basic statistical implications of interpolation in the complex domain (Section 2) and the intensity domain (Section 3). The treatment used is simple in the sense that we only deal with preservation of speckle statistics in untextured areas. The implications for measurements of texture are discussed at a number of points in the sequel.

2. Resampling the complex signal.

In this section we examine the statistical implications of performing an interpolation of the form

$$m_k = \sum_{j=-\infty}^{\infty} c_j n_{k+j} \quad (4)$$

on the complex data $\{n_k\}$, before forming the intensity image $|m_k|^2$ where the c_j are constants and we in fact expect $c_j = 0$ unless $N_0 < j < N_1$, some N_0, N_1 (i.e. a finite interpolation algorithm).

The interpolation (4) is simply a linear filtering operation. Over homogeneous areas the n_k are wide-sense stationary circular Gaussian mean-zero random variables with variance σ in the real and imaginary channels, and uncorrelated real and imaginary parts. Hence the m_k will be Gaussian (Papoulis, 1984) and zero-mean, and we can show readily that the real and imaginary parts of the m_k are uncorrelated with standard deviation $\sigma\sqrt{E}$, where

$$E = \sum_{k=-\infty}^{\infty} |c_k|^2 \quad (5)$$

The intensity image $|m_k|^2$ will therefore have negative exponential statistics, as desired, with mean value $2\sigma^2 E$. Preservation of the mean therefore gives the constraint

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = 1 \quad (6)$$

which was violated in the AgriSAR data.

As well as the single point statistics, it is important to consider the two point statistics resulting from (4). This is because textural information is present in SAR data and can be found using autocorrelation concepts (Oliver, 1989). We consider here only the autocorrelation function (acf) of the complex and intensity data, and show that interpolation introduces correlation which must not be interpreted as texture.

Direct calculation or use of linear system concepts shows that

$$\begin{aligned} R_m(j) &= E(m_{k+j} m_k^*) \\ &= \sum_k \sum_l c_k c_l^* R_n(j+k-l) \end{aligned} \quad (7)$$

where $R_m(j)$ is the acf of $\{m_k\}$, j is an integer, E is expectation and $*$ denotes complex conjugate. $R_n(j)$ is defined similarly. In the ideal Nyquist sampled case

$$R_n(j) = \begin{cases} 2\sigma^2 & j = 0 \\ 0 & j \neq 0 \end{cases} \quad (8)$$

and hence

$$\begin{aligned} R_m(j) &= 2\sigma^2 \sum_k c_k c_{k+j}^* \\ &= 2\sigma^2 \rho_h(j) \end{aligned} \quad (9)$$

where $\rho_h(j)$ is the acf of the system function $[h]$ associated with (4), and we have used (6).

We note in passing that no finite set of $\{c_k\}$ can permit $R_m(j) = R_n(j)$ for all j , except the trivial case $c_k = \pm 1$ some k , $c_j = 0$ for $j \neq k$, which is equivalent to translating the whole data set. (An infinite set $\{c_k\}$ allowing $R_m(j) = R_n(j)$ is afforded by $\{\frac{\sin \pi(k-\epsilon)}{\pi(k-\epsilon)}\}$ for any ϵ , but other solutions are possible). Hence the auto-correlation structure of the data must be changed by this operation.

Upon forming the intensity image $I_k = |m_k|^2$, we can use the Siegert relationship to show that

$$\begin{aligned}
 C_I(j) &= E(I_{k+j} I_k) - E^2(I_k) \\
 &= 4\sigma^4 |p_h(j)|^2 \\
 &= 4\sigma^4 \left| \sum_k c_k c_{k+j}^* \right|^2 \quad (10)
 \end{aligned}$$

Note that the variance of I is given by

$$\sigma_I^2 = C_I(0) = 4\sigma^4 \quad (11)$$

using (6). Equation (10) describes the effects of any interpolation scheme based on the complex data. However, it is not complete because it treats only the autocovariance structure of the interpolated samples. In many cases in going from the complex to the intensity image we want to insert extra samples, mainly because of bandwidth and Nyquist sampling considerations. We deal here only with the simplest case where we insert one extra point between each pair of complex sample values using equation (4) before forming the intensity image. We assume that the intensity value corresponding to the original data value n_k is $|n_k|^2$. Hence the intensity data $\{I_k\}$ are given by

$$I_k = \begin{cases} |n_{\lambda}|^2 & k = 2\lambda \\ |m_{\lambda}|^2 & k = 2\lambda + 1 \end{cases} \quad (12)$$

where m_{λ} is given by (4).

Then, for the ideal case given by (8), the auto covariance is no longer stationary. In particular,

$$\begin{aligned}
 \frac{1}{4\sigma^4} C_I(p, q) &= \frac{1}{4\sigma^4} E(I_{k+p} I_{k+q}) - 1 \\
 &= \begin{cases} 1 & p = q \\ 0 & k+p \text{ \& } k+q \text{ even, } p \neq q \\ |p_h(\frac{p-q}{2})|^2 & k+q \text{ \& } k+p \text{ odd, } p \neq q \\ \left| \frac{c_{\frac{p-q}{2}} + 1}{2} \right|^2 & k+p \text{ even, } k+q \text{ odd} \\ \left| \frac{c_{\frac{q-p}{2}} + 1}{2} \right|^2 & k+p \text{ odd, } k+q \text{ even} \end{cases} \quad (13)
 \end{aligned}$$

where we have used (11). This complicated acf structure must be borne in mind when interpreting any apparent textural properties of the data. Note that none of these problems occur if the original complex data is oversampled. Then the intensity image corresponds to genuine signal values and stationarity is guaranteed.

3. Resampling intensity data

In this section we assume that an intensity image $\{I_k\}$ has been formed from an oversampled complex image. If the complex image has been formed by interpolation, the results of the previous section tell us that the situation is more complicated. The intensity image has a covariance which is not simply a delta function, because of the oversampling. (The form of this autocovariance is readily defined in terms of the complex correlation function of the SAR, using the Siegert relation again (Quegan et al, 1986)).

For a resampling scheme similar to (4)

$$p_k = \sum_j d_j I_{k+j} \quad (14)$$

where the d_j are constants, we again have a linear filtering operation. However, we can no longer appeal to the helpful properties of Gaussian inputs, since the $\{I_k\}$ are correlated exponentially distributed random variables. As a result, we will no longer preserve the probability distribution function; $\{p_k\}$ will not be exponentially distributed. Establishing the form of the pdf of $\{p_k\}$ in the presence of correlation is complicated and is not discussed here. However, we can illustrate that changes occur by considering the unrealistic case of independent exponentially distributed $\{I_k\}$. Then the characteristic function of p_k is given by

$$\Phi_p(\omega) = \prod_j \Phi_I(d_j \omega) \quad (15)$$

where $\Phi_I(\omega)$ is the characteristic function of the random variable I_k , and since I_k is exponentially distributed with mean $2\sigma^2$ has the form

$$\Phi_I(\omega) = \frac{1}{1 - 2i\sigma^2\omega} \quad (16)$$

Without going into details, it is clear that the product on the right-hand side of (14) can be expanded into partial fractions. If all the d_j are distinct this will give rise to a weighted sum of the form

$$\Phi_p(\omega) = \sum_j A_j \Phi_I(d_j \omega) \quad (17)$$

where the A_j are constants, and the pdf of $\{p_k\}$ can be found by an inverse transform (which will yield a weighted sum of negative exponentials). If some of the d_j are repeated, this will give rise to gamma

function terms.

It is clear that, in general, $\{p_k\}$ and $\{I_k\}$ have different distributions (different characteristic functions), but we note two exceptions. The first is the trivial case, where all the d_j are 0 except for some $d_k = 1$. The second is when the set of d_j is infinite and corresponds to Nyquist samples of a $\sin x/x$ function. Similar conclusions apply in the presence of correlation.

Since the pdf is not preserved in general, we may choose to impose constraints to preserve other quantities. In particular, preservation of the mean implies

$$\sum_j d_j = 1 \quad (18)$$

Since $R_p(k) = E(p_{j+k} p_j)$

$$= \sum_j \sum_{j'} d_j d_{j'} R_I(k+j-j') \quad (19)$$

preservation of the variance implies

$$8\sigma^4 = R_p(0) = \sum_j \sum_{j'} d_j d_{j'} R_I(j-j') \quad (20)$$

Preservation of each moment or successive lag of the acf yields a new constraint. Solutions to these constraints are still being investigated.

4. Conclusions

- (a) Interpolation of SAR data in the complex image yields correct intensity distributions and mean preservation if constraint (7) is met. The autocorrelation structure of the intensity data becomes non-stationary and complicated unless the complex data is already over-sampled, or an infinite interpolation scheme is used.
- (b) Interpolation in the intensity image affects both the probability distributions and the autocorrelation structure. A full treatment taking account of correlation of samples is currently being developed.
- (c) Further work is required to investigate the effects of resampling on textural measurements.

5. References

- Clark C, Hendry A & Sowter A (1989), Resampling and SAR speckle statistics (to be submitted to Int J Remote Sensing).
- Freeman A and Keen A (1986), Note on aspects of VARAN-S SAR image quality, GEC Internal Rpt Y/218/4105.
- Oliver C J (1989), Natural clutter in coherent imaging (submitted to Int J Remote Sensing).
- Papoulis A (1977), Signal Analysis, McGraw Hill.
- Papoulis A (1984), Probability, Random Variables and Stochastic Processes, McGraw Hill.
- Quegan S, Lamont J M, Miller R J, Meadows P J, Wright A and Veck N J (1986), The simulation of SAR data products, ESA SP-264, 37.
- Quegan S and Yanasse C (1989), Detection of changes in the backscatter from agricultural plants using AgriSAR 86 data, Final Rpt to JRC, Ispra.

Minimum Mean-Square Error Resampling for Remotely Sensed Imagery

Gregory W. Wornell, David S. Kauffman, Bruce Sharpe

MacDonald Dettwiler & Associates Ltd.
13800 Commerce Parkway, Richmond, B.C.
Canada V6V 2J3

Abstract

Two families of optimum realizable resampling kernels are developed for use with remotely-sensed data in clean and noisy environments. These robust, model-based kernels are designed for minimum mean square resampling error.

The first family of kernels is designed for use with Nyquist-sampled imagery. The resulting kernels offer performance as good as the best of the popular resamplers in clean environments, and superior performance in noisy environments.

The second family of kernels is designed for resampling aliased imagery. For smaller kernels, resampling performance improvement is achieved over traditional resamplers. Most notably, experiments show that the 4×4 -point kernel can reduce resampling error by more than 22% over cubic convolution.

Introduction

The need for radiometrically-accurate resampling algorithms arises regularly in problems of quantitative remote sensing. Indeed, resampling plays a crucial role in the correction of raw satellite image data and its transformation to a UTM coordinate grid.

In general, the problem of resampling or interpolation involves the reconstruction of a waveform known only at a finite collection of points—usually on a regular sampling grid. Most interpolators currently in use fall into two categories:

Spline-based Interpolation Here, the sampled data is splined together with curves of the desired form. Nearest-neighbour, bilinear, and cubic-convolution interpolation are examples of this method [3], corresponding, respectively, to piecewise-constant, linear, and cubic-polynomial splines.

Sinc-based Interpolation Here it is assumed that the waveform to be reconstructed is band-limited and that the sampling grid is sufficiently dense. In this case, optimum interpolation requires the use of sinc functions and is, in fact, *unrealizable*. Realizable approximations to this method have relied on the use of windowing techniques [4] [3].

Inherently, both approaches make use of certain smoothness assumptions about the waveform being reconstructed¹. And, while the resulting interpolators have been shown empirically effective both subjectively and in terms of RMS error—they have not been designed

to be optimal in any particular sense.

In this work, we describe a framework in which to design families of optimum realizable resampling kernels which give minimum mean-square resampling error based on some simple but reasonable models for the underlying imagery. Some related formulations have been considered in [1] [2].

A Framework

Through the remote sensing process, distorted and corrupted samples of the ground radiance image constitute the available raw data. While various effects such as Earth curvature and satellite motion cause these samples to be irregularly spaced, we shall assume that the sampling falls approximately on a rectangular lattice at some rotation.

It will also be convenient to employ imagery models with a degree of statistical *separability* with respect to the along-scan and across-scan directions. While this assumption has no strong physical basis, it ensures that we obtain separable resampling kernels suitable for two-pass processing, and it allows us to consider along-scan and across-scan resampling as two separate one-dimensional problems.

In its discrete form, the general one-dimensional resampling environment is as shown in Fig. 1. The continuous-time signal $z(t)$ represents the radiance of ground imagery. The linear shift-invariant filter $g(t)$ models the sensor distortions before and during sampling. With Landsat data, for example, $g(t)$ can be used to account for detector size effects (cross-scan) and band-pass filtering (along-scan) in the sensor as well as non-ideal sampling. The additive sequence $w[n]$ models noise in the acquisition process arising during sensing, sampling at rate T , and quantization. The available data for resampling is then $v[n]$.

The resampler is represented by the filter $\hat{h}[n]$. To ensure that this filter is realizable and stable, we enforce the constraint that $\hat{h}[n]$ have finite extent, i.e.,

$$\hat{h}[n] = 0, \quad n < -M, \quad n > N. \quad (1)$$

for some $M < N$ (usually $M = -N + 1$).

Finally, we want the resampler output at some instant n_0 to best recover $z(t_0)$ for an arbitrary t_0 from the sequence $v[n]$, though we recognize that during subsequent stages of processing we may ultimately be interested in recovering $z(t_0)$.

¹Indeed, without some additional information there is no "correct" way to reconstruct the waveform.

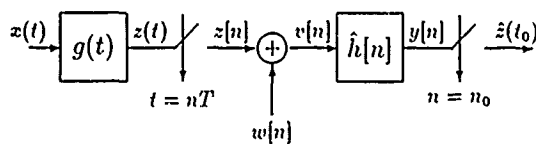


Figure 1: The general resampling environment.

We shall assume that $x(t)$ is a zero mean, wide-sense stationary random process. The zero-mean assumption is not serious, for we may always process our non-zero mean imagery by removing the mean prior to resampling and restoring it immediately thereafter. The wide-sense stationarity assumption is somewhat stronger, but results in resampling kernels which are invariant over the imagery. As a consequence of stationarity, we may write the autocorrelation of $x(t)$ as

$$R_x(\tau) = E\{x(t)x'(t - \tau)\} \quad (2)$$

The noise sequence $w[n]$ is considered to be zero-mean and white with power σ^2 , so that

$$R_w[k] = E\{w[n]w[n - k]\} = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad (3)$$

Kernels for Band-limited Imagery

In this section, we assume that $x(t)$ is strictly band-limited and that the corresponding $z(t)$ has been sampled at or above the Nyquist-rate. Provided the sampling is noise-free ($\sigma^2 = 0$) it is well-known that $z(t_0)$ may be recovered, in principle, from the samples

$$v[n] = z[n] = z(nT), \quad n = 0, \pm 1, \pm 2, \dots \quad (4)$$

In particular, choosing a suitable $\alpha \in [0, 1)$ and integer n_0 such that

$$t_0 = (n_0 + \alpha)T, \quad (5)$$

Whittaker-Shannon sampling theory assures us that

$$z(t_0) = (h * z)[n_0] \quad (6)$$

where

$$h[n] = \text{sinc}(n + \alpha) = \frac{\sin \pi(n + \alpha)}{\pi(n + \alpha)}. \quad (7)$$

Unfortunately, this sinc filter, corresponding to band-limited interpolation, has infinite extent and is, in fact, unrealizable. Moreover, we have only a finite segment of the sample sequence $z[n]$ available. While the truncated sinc filter seems a natural candidate for our resampling kernel, we shall see that, in fact, it is optimal only for a rather specific and atypical scenario.

Using a Wiener filtering framework, we may define our objective as minimizing the mean-square resampling error

$$\epsilon = E\{[z(t_0) - \hat{z}(t_0)]^2\} \quad (8)$$

where $\hat{z}(t_0)$ is the output of filter $\hat{h}[n]$ at n_0 , viz.,

$$\hat{z}(t_0) = y[n_0] = \sum_{k=M}^N \hat{h}[k]z[n_0 - k] + w[n_0 - k]. \quad (9)$$

Note that we may express $z(t_0)$ in terms of the ideal interpolator $h[n]$ by

$$z(t_0) = \sum_{k=-\infty}^{\infty} h[k]z[n_0 - k]. \quad (10)$$

By minimizing (8) using (9) and (10), it is straightforward to show that the optimal resampler $\hat{h}[n]$ satisfies the Toeplitz normal equations

$$\sum_{k=M}^N \hat{h}[k]R_z[n - k] + \hat{h}[n]\sigma^2 = \sum_{k=-\infty}^{\infty} h[k]R_z[n - k] \quad (11)$$

for $n = M, M + 1, \dots, N$.

If the samples $z[n]$ are modeled as a first-order autoregressive (AR(1)) sequence, then

$$R_z[k] \propto \rho^{|k|} \quad (12)$$

for some $|\rho| < 1$. These are popular simplified models for image processing. Note that ρ is a measure of the degree of correlation in the data, with $\rho = 0$ corresponding to white (uncorrelated) data. Empirical studies have shown that, for a variety of imagery, values of ρ near 1 are most applicable.

Using the AR(1) model, the normal equations may be solved to give the optimum resampler

$$\hat{h}[n] = \begin{cases} \sum_{k=0}^{\infty} h[-k + M]\rho^k & n = M \\ h[n] & M < n < N \\ \sum_{k=0}^{\infty} h[k + N]\rho^k & n = N \end{cases} \quad (13)$$

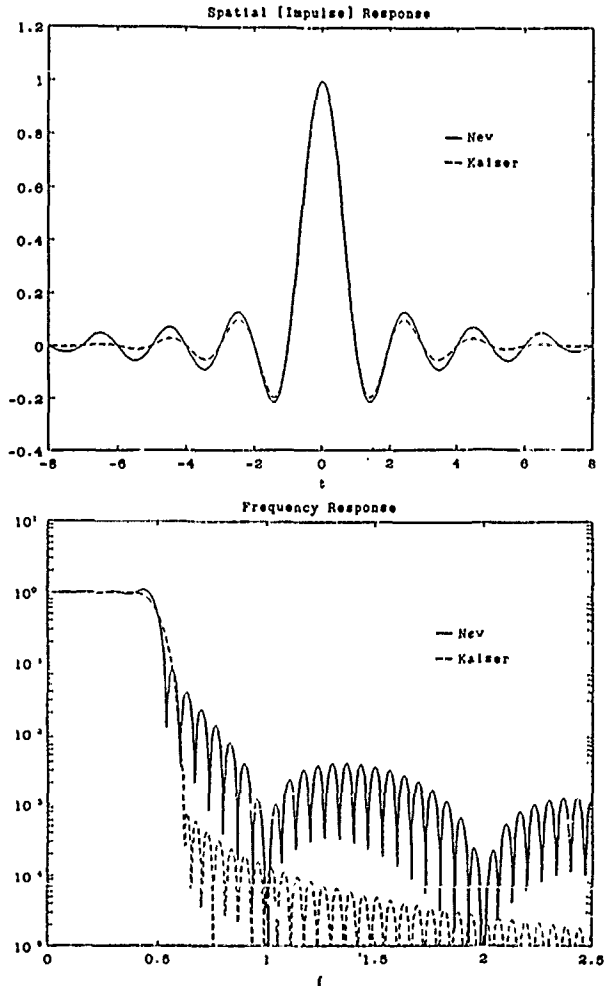


Figure 2: Spatial and frequency response characteristics of the optimal 16×16 -point interpolator for band-limited imagery corresponding to $\rho \sim 1$ and $\sigma^2 = 0$. The Kaiser-windowed sinc interpolator is shown for comparison.

Evidently, this resampler is a generalization of the truncated-sinc filter, for when $\rho = 0$, $\hat{h}[n]$ is indeed a truncated sinc. However, for ρ near 1, $\hat{h}[n]$ is a truncated sinc whose end-points are reduced in magnitude. Note that in the limiting case $\rho \rightarrow 1$, the filter is normalized

$$\sum_{n=M}^N \hat{h}[n] = \sum_{k=-\infty}^{\infty} h[k] = 1, \quad (14)$$

obviating the need for mean-removal in the resampling process. Fig. 2 compares the spatial and frequency characteristics of the 16×16 -point interpolator corresponding to $\rho \sim 1$ and $\sigma^2 = 0$ with that of the Kaiser-windowed sinc interpolator.

| Kernel | | | Noise | Error (DN) | |
|----------|----------|----------------|-------|------------|------|
| Type | ρ | size | SNR | RMS | Peak |
| New | ~ 1 | 2×2 | - | 4.9 | 25 |
| Bilinear | - | 2×2 | - | 5.1 | 26 |
| New | ~ 1 | 4×4 | - | 3.8 | 17 |
| Cubic | - | 4×4 | - | 3.2 | 15 |
| New | ~ 1 | 8×8 | - | 2.0 | 11 |
| Hanning | - | 8×8 | - | 1.8 | 8 |
| New | ~ 1 | 16×16 | - | 1.6 | 8 |
| Kaiser | - | 16×16 | - | 1.4 | 7 |
| New | 0.9 | 16×16 | 1 dB | 16.5 | 75 |
| Kaiser | - | 16×16 | 1 dB | 27.0 | 95 |
| New | 0.9 | 16×16 | 11 dB | 7.4 | 36 |
| Kaiser | - | 16×16 | 11 dB | 8.8 | 36 |

Table 1: Resampling performance of kernels for band-limited imagery.

Resampling Experiments

To evaluate the resampling kernels derived in this section, a 128×128 -pixel test chip was twice resampled by one-half pixel both horizontally and vertically, and compared to a displaced version of the original chip. In some tests, the chip was artificially degraded by additive white Gaussian noise to achieve a prescribed signal-to-noise ratio (SNR) prior to resampling. As appropriate, comparisons are made to the traditional resampling schemes: bilinear interpolation, cubic convolution, Hanning windowed sinc interpolation, and Kaiser-windowed ($\beta = 6$) sinc interpolation. As the results in Table 1 indicate, the new kernels are competitive with the best of the traditional resamplers in noise-free environments despite a markedly different kernel shape. Moreover, in the presence of noise, the new kernels offer much improved performance.

Kernels for Aliased Imagery

In the last section, it was assumed that $x(t)$ was band-limited. Unfortunately, in practice, imagery is only approximately band-limited and, in any case, is always sampled in a manner which introduces some aliasing. In this section, we develop an alternative set of kernels based upon a model which accounts for aliasing introduced in the sensing process.

We begin by modeling $x(t)$ as a continuous AR(1) process for which the autocorrelation function satisfies

$$R(\tau) = E[x(t)x(t-\tau)] \propto e^{-\omega_0|\tau|} \quad (15)$$

where $\omega_0 > 0$ is a parameter. The choice of ω_0 specifies the degree of correlation in the data, and values near 0 are appropriate for imagery. Since the power spectrum for this process is

$$S_x(\omega) = \mathcal{F}\{R(\tau)\} \propto \frac{1}{1 + (\omega/\omega_0)^2}, \quad (16)$$

it is evident that $x(t)$ is, indeed, not band-limited. Note, also, that for some $0 < \rho < 1$, (15) may be re-written as

$$R(\tau) \propto \rho^{|\tau|/T} \quad (17)$$

where, evidently, $\omega_0 \approx 0$ corresponds to $\rho \approx 1$.

Now for arbitrary $R_x(\tau)$, the minimization of (8) subject to (9) gives the optimal resampler as the solution to the normal equations

$$\sum_{k=M}^N \hat{h}[k] R_x((n-k)T) + \sigma^2 \hat{h}[n] = R_x((n+\alpha)T) \quad (18)$$

for $n = M, M+1, \dots, N$.

In the absence of $g(t)$, this optimum resampler is given by

$$\hat{h}[n] = \begin{cases} (\rho^{1(1-\alpha)} - \rho^{(1-\alpha)})/(\rho^{-1} - \rho) & n = 0 \\ (\rho^{-\alpha} - \rho^{\alpha})/(\rho^{-1} - \rho) & n = 1 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

which gives bilinear interpolation in the limit $\rho \rightarrow 1$. This rather surprising result, also deduced by Polydoros and Protonotarios [2], establishes that for such highly-correlated signals a two-point resampler is optimal. In fact, this is a consequence of the extensive aliasing in the samples of these signals.

More realistically, some low-pass filtering takes place in the sensor. In fact, if the sensor performs natural sampling, so that $g(t)$ is given by

$$g(t) = \begin{cases} 1/T & |t| < T/2 \\ 0 & |t| \geq T/2 \end{cases}, \quad (20)$$

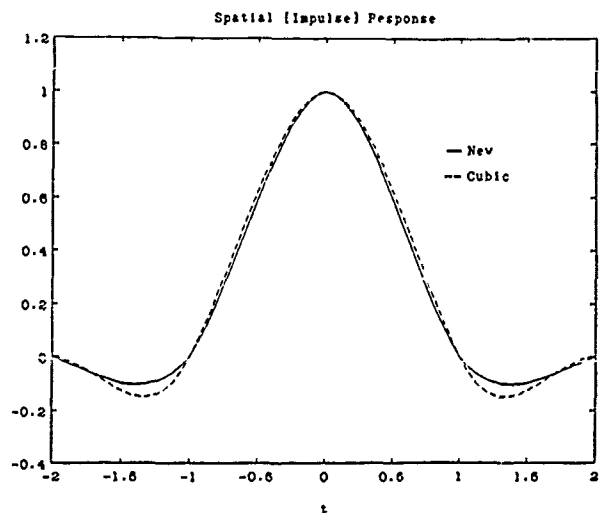
then

$$R_z(\gamma T) = \begin{cases} [\rho^{1+|\gamma|} - 2\rho^{|\gamma|} + \rho^{1-|\gamma|} + 2(|\gamma|-1)\mu]/\mu^2 & |\gamma| \leq 1 \\ [\rho^{|\gamma|+1} - 2\rho^{|\gamma|} + \rho^{|\gamma|-1}]/\mu^2 & |\gamma| > 1 \end{cases} \quad (21)$$

where

$$\mu = \ln |\rho|. \quad (22)$$

Using this expression for $R_z(\cdot)$, the normal equations (18) may be solved numerically to obtain resamplers parameterized by ρ and σ^2 . In noise-free environments ($\sigma^2 = 0$), these resamplers have a damped sinc appearance similar to be that of the Kaiser-windowed sinc interpolators, but with stronger damping. Fig. 3 compares the spatial and frequency characteristics of the 4×4 -point interpolator corresponding to $\rho = 0.9$ and $\sigma^2 = 0$ with those of cubic convolution.



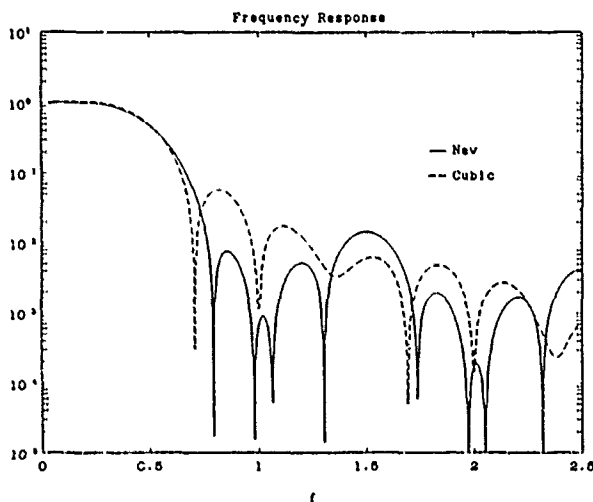


Figure 3: Spatial and frequency response characteristics of the optimal 4×4 -point interpolator for aliased imagery corresponding to $\rho = 0.9$ and $\sigma^2 = 0$. The cubic convolution interpolator is shown for comparison.

| Kernel | | | Error (DN) | |
|----------|--------|----------------|------------|------|
| Type | ρ | size | RMS | Peak |
| New | 0.9 | 2×2 | 4.9 | 24 |
| Bilinear | - | 2×2 | 5.1 | 26 |
| New | 0.9 | 4×4 | 2.5 | 12 |
| Cubic | - | 4×4 | 3.2 | 15 |
| New | 0.9 | 8×8 | 2.1 | 10 |
| Hamming | - | 8×8 | 1.8 | 8 |
| New | 0.9 | 16×16 | 2.1 | 10 |
| Kaiser | - | 16×16 | 1.4 | 7 |

Table 2: Resampling performance of kernels for aliased imagery.

Resampling Experiments

An experimental framework similar to that described earlier was employed in the evaluation of these resampling kernels, though only noise-free ($\sigma^2 = 0$) scenarios were considered this time. Table 2 indicates the resampling results using kernels corresponding to $\rho = 0.9$. (In fact, resampling performance was observed to be relatively independent of ρ .) Improvements in performance can be noted for the smaller sized resamplers—indeed, the 4×4 -point resampler gives 22% less resampling error than cubic convolution.

Discussion

Despite the simplicity of the models employed in this work, some remarkably effective resampling kernels have been developed for use with remotely-sensed data.

The first family, based on an overly-optimistic model assuming no aliasing in the sensing process, yields kernels whose performance virtually indistinguishable from the best of the popular resamplers, particularly in the case of large kernels. However, the new kernels, though sinc-like in appearance, exhibit considerably less damping than traditional Kaiser-windowed interpolators. In noisy environments, the appropriate optimum kernels significantly outperform the traditional resamplers, optimally combining noise removal (smoothing) and interpolation. These kernels may be of use in some resampling scenarios.

The second family of kernels is based on an overly-pessimistic model which assumes that significant aliasing is introduced in the sensing process. These kernels are especially effective for low-order interpolation, such as in the case of 4×4 -point resampling. Much of this gain is due to the fact that the false contrast enhancement of cubic convolution is avoided. Kernels of this family also have a damped-sinc appearance, though the degree of damping is more pronounced than is commonly introduced with, for example, typical Kaiser windows.

Note that while it may seem tempting to by-pass the autoregressive autocorrelation models $R_z(\cdot)$ of this work and estimate the required values directly from data, this approach has some inherent difficulties:

1. it is only possible to estimate *samples* of the autocorrelation from our sampled data;
2. the quality of the estimation may be poor if the σ^2 is large; and
3. some additional assumption about the data such as band-limiting must be made to enable interpolation of the autocorrelation. This effectively prohibits the incorporation of aliasing into the framework.

Nevertheless, for the noise-free, band-limited case, this approach is feasible and was considered, in fact, by Shlien [3]. However, while kernels adapted locally to the data can give excellent resampling error performance, the results can be misleading. After accounting for the additional computation inherent in the accumulation of statistics, one finds, in practice, that the same level of performance can be attained much more cheaply simply by using a larger sized resampling kernel.

In conclusion, therefore, we have described an apparently useful framework for the design of interpolators for remotely-sensed data. Future work ought to address the development of more refined models for both the imagery and the satellite-dependent sensing process.

References

- [1] G. Oetken, T.W. Parks, and H.W. Schüssler, "New Results in the Design of Digital Interpolators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, no. 3, pp. 301-309, June 1975.
- [2] A.D. Polydoros and E.N. Protonotarios, "Digital Interpolation of Stochastic Signals," *IEEE Trans. Circuits and Systems*, vol. CAS-26, no. 11, pp. 916-922, Nov. 1979.
- [3] S. Shlien, "Geometric Correction, Registration, and Resampling of Landsat Imagery," *Canadian J. Remote Sensing*, no. 5, pp. 74-89, 1979.
- [4] A.V. Oppenheim and R.W. Schaffer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

EFFECT OF RESAMPLING ON THE GEOMETRIC AND RADIOMETRIC FIDELITY OF DIGITAL IMAGES

Eugene E. Derenyi and Raad K. Saleh

Department of Surveying Engineering
University of New Brunswick
Fredericton N.B. Canada E3B 5A3

ABSTRACT

A Landsat Thematic Mapper image was used to study the effect of resampling, by nearest neighbour (NN) and by cubic convolution (CC) interpolation, on the outcome of subsequent image classification. The image was subjected to 9° and 35° rotations before resampling. The analysis was based on comparing pixel counts in classes obtained from the original and resampled images and on examining the confusion matrices. It has been concluded, that resampling can alter the radiometric content of images to an extent that pixel-by-pixel agreement between the classification of the original and resampled data is only in the order of 60% to 80%.

Keywords: classification accuracy, geometric fidelity, radiometric fidelity, registration, resampling.

INTRODUCTION

Numerous applications require that digital images be fitted to a geographic reference system, or matched with another image. This operation is performed by two transformations, a geometric and a radiometric. The first transformation calculates new pixel locations with respect to a reference system or another image. The latter involves the generating of new radiometric value for each new pixel by a mathematical interpolation, called resampling. Several resampling techniques have been developed with different characteristics in regard to geometric accuracy, radiometric fidelity, computing time, etc. These techniques are well documented in the literature.

The accuracy, with which a resampling technique reconstructs the original spatial and radiometric data, must be known, in order to assess the impact of resampling on the subsequent application. Each time an image is resampled, the radiometric values, represented by the digital numbers (DN) of pixels, are modified, changing the original information content of the image. Consequently, this influences the subsequent visual, statistical, and quantitative analyses. As Logan and Srahler [1979] has stated:

Use of one resampling algorithm over another will produce a different classification result because different algorithms generate different density values.

Several studies, pertaining to this problem, have been undertaken, but still no consistent findings have emerged, as echoed by the following statements made in the scientific literature:

Little research work has been done to quantify the error introduced by each algorithm or suggest when to use one

technique over another. [Logan and Srahler, 1979].

Although it is well recognized that different resampling algorithms have different effects on digital values of the output images, little work has been done to assess the implications of this on classifier effectiveness. [Belward and Taylor, 1986].

Although much research has been performed on resampling techniques, it is evident that the general effect of resampling prior to classification on map accuracy are [sic] not known. Few other disciplines allow one to manipulate data values to analysis, without at least knowing the consequences of these manipulations upon the results. [Smith and Kovalick, 1985].

The research reported in this paper is an attempt to shed some light on this problem, through experimentation with images.

EXPERIMENT AND RESULTS

Training Area Selection

A Landsat TM subscene of 172 lines by 313 pixels was selected for this study. It contained different land cover types including agricultural land, urban area, forest, and water. Two training sites were selected, one representing agricultural fields and the other water bodies.

The image, on which the training areas were identified by polygons, was rotated by 9° and 35°. Each of the rotated images was then resampled by nearest neighbour (NN) and cubic convolution (CC) interpolation without changing the pixel size. The 9° rotation corresponds to the orbital inclination of earth observation satellites, while the 35° represents large rotations which could occur when airborne data are geographically registered.

The training area polygons were rotated by the same angles, thus assuring that the polygons in each of the rotated and resampled images represented the same physical, i.e., geographical, location as in the original image. Spectral signatures were then generated for the two classes in the original image as well as in each of the four resampled images. The training area statistics for Band 5 are shown in Table 1.

This table shows that resampling affected the DNs of pixels within the same physical area in a varying degree, depending on the resampling technique and rotation angle employed. While the NN caused minimal changes in the mean and standard deviation (SD), CC has noticeably shifted the mean and increased the SD.

Moreover, the size of the rotation angle and the original value of the SD had an apparent effect on the dispersion of DNs.

Next, the two land cover types were separated in the entire test image by a maximum likelihood classifier (ML), to ascertain the effect of resampling on image classification.

Classification

The usual practice, in classification, is to use all or several of the bands in a multispectral data set. In this experiment, however, only a single band was used since the objective was to study the changes in the classification results rather than to generate accurate quantitative information. Band 5 was selected because both classes were easily distinguishable.

The two factors considered in the implementation of the ML classification were:

- the scheduling of the training area selection, whether it has been done before or after resampling, and
- the a priori probability level selected.

Two cases of classification with various probability levels were used to test the above factors. In the first case, the signatures were generated from the original image, whereas in the second case the signatures were taken from the resampled image.

Case 1, Original Signatures

The signatures used in this case were generated from the original, unresampled image. These signatures were, however, used for the classification of the original image, as well as for the four rotated and resampled images. Probability levels of 65%, 80% and 95% were used. The classification results, shown in Table 2, were represented by pixel counts.

With few exceptions, no significant changes existed in the classification results between the original and the resampled images. ML classification at the 95% probability level provided the best agreement. The largest discrepancy encountered was in the order of 7%. It appeared therefore, that neither the resampling technique, the probability level, nor the rotation angle had a significant effect on classification results, in terms of pixel counts.

This table, however, provided only overall pixel counts and did not indicate any trade off that may have occurred between the two classes or their surroundings. In other words, it was not known, whether the shape and size of individual class clusters were the same in the original and in the resampled images. A confusion matrix was therefore generated to compare the pixel counts of the classes, in the NN and the CC resampled images. This comparison was only performed on classes generated at the 95% probability level, where the pixel count of the original and the resampled image classes had the best agreement. The confusion matrices are shown in Table 3.

This table showed a significant discrepancy in the classification of individual pixels of the NN and CC resampled images. The level of agreement ranged from 73% to 79%. At 9° rotation for instance, only 564 pixels out of the total of 770 were assigned to Class 1 in both cases, which is a 73% agreement. While 206 pixels were included in the NN image as members of that class, they were excluded from the same class in the CC image and were left unclassified. At the same time, 196 pixels of the CC image in Class 1 were excluded from that class in the NN resampled images. Nevertheless, the total pixel count for Class 1 was 770 and 760 in the NN and CC resampled images respectively. This was an excellent agreement for providing reliable class area estimation within a certain geographic boundary. Serious problems would arise, however, when area estimates would be required for individual class clusters, such as the area of specific

water bodies or farm fields.

Case 2, Resampled Signatures

In this case, the signatures were generated from the training areas in the resampled images, called hereinafter 'resampled' signatures. The training statistics were shown in Table 1. Each signature was used in classifying the same image, from which it was generated. For instance, in classifying the image which was rotated by 9° and resampled by CC, a mean DN of 151.25 and SD of 6.88 was used for Class 1, and a mean DN of 7.47 and SD of 2.11 for Class 2. The classification results were listed in Table 4. Upon examination of this table, the following observations were made:

- Good agreement existed between the pixel counts for the classification of the original image and the one resampled by NN if the rotation angle was small. Significant discrepancies occurred, however, when the rotation angle was large.
- Large discrepancies were apparent in the pixel counts if the resampling was performed by CC, regardless of the size of the rotation angle.
- There was some evidence that the discrepancies became larger with an increase in the probability level.
- A comparison of Tables 2 and 4 indicated that the results obtained with the resampled signatures were inferior to those generated with the original signatures.
- All discrepancies of significant magnitude in the pixel count of the resampled versus the original image classification were positive.

The latter statement was underlined by the confusion matrix, which compared two classifications of CC resampled images, i.e., one with original signatures and the other with resampled signatures, (see Table 5). This table showed that a class formed with resampled signatures included all the pixels which were classified as the same category, using original signature. Additional pixels were, however, included in this class, when the resampled signature were used. This meant that signatures, generated after resampling exaggerated the class size.

Finally, classes of CC resampled images, which were obtained using resampled signatures, were compared with classes of NN resampled images, using original signatures. The results were listed in Table 6. It was apparent, that the percentage of pixels classified the same in both cases, has been dramatically reduced.

For 9° rotation, the pixel count of the CC classification has matched the NN classification at the 74% level in Class 1 and at 79% in Class 2, when original signature were used, as showed in Table 3a. For the same rotation angle, these percentages were reduced to 65% and 73%, respectively, when resampled signatures were employed, as shown in Table 6a. The agreement was further reduced with an increase in the rotation angles, as shown in Table 6b.

SUMMARY AND CONCLUSIONS

Quantitative analyses have been performed in this study, to evaluate the effects of geometric registration and subsequent resampling on the spatial and radiometric properties of digital images. The evaluation has involved the comparison of resampled versus unresampled images and NN versus CC resampling.

This experiment has shown, that resampling had a profound effect on digital images. It modified the internal spatial structure and radiometric content. While NN resampling did not introduce new radiometric values, it was demonstrated, that CC had significantly influenced the image statistics, which, in turn, caused substantial changes in the outcome of subsequent quantitative analyses, in comparison with the results obtained with original, unresampled data. It is important to note the following:

- Changes in the training statistics and in the outcome of the classification were amplified by an increase of the rotation angle.
- The net effect of CC resampling was an increase in the size of classes in terms of pixel count and an alteration in the boundary and shape of individual class clusters.
- While the increase in the magnitude of a class was minimized by performing the classification with signatures derived from the original, unresampled data, the discrepancies in the extent and shape of individual class clusters remained significant.

Although, the study reported here has been performed on a rather small data set, it has, nevertheless, confirmed the conclusions reached by others, such as Verdin [1983] and Atkinson [1985], that resampling should be treated with caution. Further investigation, with large data sets and more thematic classes is in order to put this problem to rest.

ACKNOWLEDGEMENT

This study was supported in part by a National Sciences and Engineering Research Council of Canada grant in aid of research.

REFERENCES

- Atkinson, P. (1985) - *The Effects of Resampling on Boundaries in Thematic Mapper Imagery*, International Conference on Advanced Technology for Monitoring and Processing Global Environmental Data, p.p. 153-162, University of London, UK, 10-12 September.
- Belward, A. S. and Taylor, J.C. (1986) - *The Influence of Resampling Method and Multitemporal LANDSAT Imagery on Crop Classification Accuracy in the United Kingdom*, Proceedings of IGARSS' 86 Symposium, p.p. 529-534, Zurich, 8-11 September.
- Logan, T. and Strahler, A. (1979) - *The Errors Associated with Digital Resampling of LANDSAT Forest Imagery for Multidate Registration*, Seventh Annual Remote Sensing of Earth Resources Conference, Vol. VIII, Tennessee, p.p. 163-179, March 27-29.
- Smith, J. and Kovalick, B. (1985) - *A Comparison of The Effects of Resampling on The Accuracy of a LANDSAT Derived Cover Type Map*, Advanced Technology for Monitoring and Processing Global Environmental Data, conference proceedings, pp.391-399, London, England, September 9-12.
- Verdin, J. (1983) - *Corrected Vs. Uncorrected LANDSAT 4 MSS Data*, LANDSAT Data Users Notes, No. 27, pp.4-8, June.

| α | Class 1 : Field | | | Class 2 : Water | | |
|----------|-----------------|---------|--------|-----------------|---------|------|
| | P.C. | Mean DN | S.D. | P.C. | Mean DN | S.D. |
| 0° | 164 | 151.40 | 5.33 | 219 | 8.10 | 1.47 |
| 9° | N.N. | 165 | 151.35 | 220 | 8.09 | 1.47 |
| | C.C. | 165 | 151.25 | 220 | 7.47 | 2.11 |
| 35° | N.N. | 164 | 151.50 | 217 | 8.20 | 1.53 |
| | C.C. | 164 | 150.76 | 217 | 7.35 | 1.93 |

α = Rotation angle, P.C. = Pixel count, SD = Standard deviation

Table 1: Training statistics of the same image, before and after resampling.

| α % | Class 1 : Field Mean DN=151.40 SD=5.33 | | | | Class 2 : Water Mean DN=8.10 SD=1.47 | | | |
|---------------|---|------|------|-----|---|------|------|------|
| | 0° | 9° | | 35° | 0° | 9° | | 35° |
| | | N.N. | C.C. | | | N.N. | C.C. | |
| 65 | 366 | 369 | 370 | 365 | 1328 | 1330 | 1394 | 1304 |
| 90 | 515 | 518 | 517 | 510 | 1328 | 1330 | 1394 | 1304 |
| 95 | 766 | 770 | 760 | 768 | 2055 | 2058 | 2041 | 2030 |

Table 2: Pixel counts of ML classification with original (unresampled) signatures, at three probability levels.

| NN original | | | CC original | | |
|-------------|-----|------|-------------|----------|---------|
| CLASS | 1 | 2 | UNCL | TOTAL CC | PERCENT |
| 1 | 564 | 0 | 196 | 760 | 74 |
| 2 | 0 | 1614 | 427 | 2041 | 79 |
| UNCL | 206 | 444 | — | — | — |
| TOTAL NN | 770 | 2058 | — | — | — |
| PERCENT | 73 | 78 | — | — | — |

a. Image rotated 9°

| NN original | | | CC original | | |
|-------------|-----|------|-------------|----------|---------|
| CLASS | 1 | 2 | UNCL | TOTAL CC | PERCENT |
| 1 | 564 | 0 | 204 | 768 | 73 |
| 2 | 0 | 1607 | 457 | 2064 | 77 |
| UNCL | 197 | 423 | — | — | — |
| TOTAL NN | 761 | 2030 | — | — | — |
| PERCENT | 74 | 79 | — | — | — |

b. Image rotated 35°

Table 3: Confusion matrix comparing ML classifications of NN and CC resampled images, at 95% probability, with original signatures.

| | | Class 1 : Field | | | | Class 2 : Water | | | |
|----|----------|-----------------|------|------|------|-----------------|------|------|------|
| % | α | 0° | | 9° | | 35° | | 0° | |
| | | N.N. | C.C. | N.N. | C.C. | N.N. | C.C. | N.N. | C.C. |
| 65 | | 366 | 369 | 495 | 365 | 551 | 1328 | 1330 | 1304 |
| 80 | | 515 | 518 | 631 | 510 | 768 | 1328 | 1330 | 1807 |
| 95 | | 766 | 770 | 938 | 781 | 1151 | 2055 | 2058 | 2354 |

Table 4: Pixel counts of ML classification with resampled signatures, at three probability levels.

| CC original | | CC resampled | | | |
|-------------|-----|--------------|-----|------|-------|
| CLASS | | 1 | 2 | UNCL | TOTAL |
| 1 | 760 | 0 | 178 | 938 | 81 |
| 2 | 0 | 2041 | 469 | 2510 | 81 |
| UNCL | 0 | 0 | — | — | — |
| TOTAL | 760 | 2041 | — | — | — |
| PERCENT | 100 | 100 | — | — | — |

a. Image rotated 9°

| CC original | | CC resampled | | | |
|-------------|-----|--------------|-----|------|-------|
| CLASS | | 1 | 2 | UNCL | TOTAL |
| 1 | 768 | 0 | 383 | 1151 | 66 |
| 2 | 0 | 2064 | 477 | 2541 | 81 |
| UNCL | 0 | 0 | — | — | — |
| TOTAL | 768 | 2064 | — | — | — |
| PERCENT | 100 | 100 | — | — | — |

b. Image rotated 35°

Table 5: Confusion matrices comparing CC resampled images, classified with original and resampled signatures, at 95% probability.

| NN original | | CC resampled | | | |
|-------------|-----|--------------|-----|------|-------|
| CLASS | | 1 | 2 | UNCL | TOTAL |
| 1 | 617 | 0 | 321 | 938 | 65 |
| 2 | 0 | 1840 | 670 | 2510 | 73 |
| UNCL | 153 | 218 | — | — | — |
| TOTAL | 770 | 2058 | — | — | — |
| PERCENT | 80 | 89 | — | — | — |

a. Image rotated 9°

| NN original | | CC resampled | | | |
|-------------|-----|--------------|-----|------|-------|
| CLASS | | 1 | 2 | UNCL | TOTAL |
| 1 | 664 | 0 | 487 | 1151 | 57 |
| 2 | 0 | 1840 | 701 | 2541 | 72 |
| UNCL | 97 | 190 | — | — | — |
| TOTAL | 761 | 2030 | — | — | — |
| PERCENT | 87 | 90 | — | — | — |

b. Image rotated 35°

Table 6: Confusion matrices, comparing NN and CC resampled images, classified with original and resampled signatures at 95% probability.

Effects of the Ephemeris Error on Effective Pointing for a Spaceborne SAR

Michael Y. Jin
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

ABSTRACT

Both Magellan SAR data acquisition and image processing require the knowledge of both the ephemeris (spacecraft position and velocity) and the radar pointing direction. Error in the knowledge of the radar pointing direction results in a loss of the signal-to-noise ratio in the image product. An error in the ephemeris data has a similar effect. To facilitate SNR performance analysis, the effect of the ephemeris error will be characterized by the effective pointing errors defined herein. This paper describes a systematic approach to relate the ephemeris error to the effective pointing errors. Result of this analysis has led to a formal accuracy requirement levied on the Magellan Navigation system.

Keywords: Ephemeris, Radar boresight, Range, Doppler, Effective Pointing Angle, and Orbit covariance Matrix

1. INTRODUCTION

In a spaceborne remote sensing system, system performance often tied to the accuracy of the ephemeris data. This is especially true for a Synthetic Aperture Radar (SAR) system such as Magellan (Venus Radar Mapper [1,2]), in which the image resolution, signal-to-noise ratio (SNR), and pixel location accuracy can be greatly affected by the ephemeris errors. An analysis to characterize this effect is very important to the planning phase of a space mission in order to levy proper requirements on the navigation system and to provide user community a set of highly confident image performance estimates.

During SAR mapping, a window with a proper delay is employed to acquire echo from ground points illuminated by a relatively narrow SAR antenna beam. An estimate for this delay is made based upon the knowledge of the spacecraft position, radar boresight pointing direction, and a planet surface model. An error in the pointing knowledge leads to an error in the delay estimate. Consequently, it results in a loss of the SNR in

the acquired echo. Similarly, this is true in the presence of a spacecraft position error. For analysis purpose, it is convenient to treat the ephemeris error similar to the pointing error by introducing a term called effective range pointing error.

In SAR signal processing (reference 3), maximal azimuth signal-to-ambiguity ratio (ASAR) can be obtained when the processing bandwidth is centered at the Doppler frequency of the intercept point of the boresight line and the planet surface. This point is also referred to as the boresight intercept point (BIP). The error in the Doppler frequency of the BIP can be caused by the error in the boresight pointing error as well as the ephemeris error. For analysis purpose, it is convenient to treat the ephemeris error similar to the pointing error by introducing a term called effective azimuth pointing error.

In the following sections, the effective pointing angles are defined. It is followed by the analysis to derive the effective pointing errors caused by the ephemeris. The performance estimate for the effective pointing errors for the Magellan mission are then presented.

2. Effective Boresight Intercept Point

In Magellan SAR processing, a range-Doppler image is formed through range and azimuth compressions. A framelet is then formed by truncating the range-Doppler image to eliminate pixels outside of the processing bandwidth and valid range swath. To get the maximal SNR for the framelet, the position of the center pixel must be in the center of the antenna beam. To achieve that purpose, the range and Doppler estimates for the boresight intercept point are made and selected as the center reference for the framelet.

In the presence of ephemeris error, the location of the center pixel of the framelet will not be in the center of the antenna beam. Instead, it will be at the effective BIP which is determined

from the true sensor position and velocity and the range and Doppler estimates. This is shown in Figure 1. Also illustrated in Figure 1 is that the estimated BIP is determined from the estimated sensor position and the pointing knowledge (which will be assumed to be correct since we will concentrate on the ephemeris error only). From the position and velocity of both the sensor and the estimated BIP, the BIP range and Doppler estimates can be made. Then, the position of the effective BIP is determined by the ground point satisfying range and Doppler constraint.

In this paper, we assume the ephemeris estimates used in SAR data acquisition and in SAR image processing are the same. Therefore, the delay estimate for setting range window and the range estimate for framelet truncation are no different.

3. Effective Radar Pointing Error

A spaceborne SAR, having the viewing geometry shown in Figure 2, consists of a curved sensor path and a target surface. A convenient coordinate system for this geometry is the local coordinate system with a radial direction \hat{r} from the planet center to the sensor, a cross-track direction \hat{c} given by the angular velocity of the sensor orbit, and an along-track direction \hat{a} given by $\hat{a} = \hat{c} \times \hat{r}$. The origin of the local coordinate system is the planet center. On the surface of either side to the track, each point is associated with a unique pair of range and Doppler values. Therefore, range and Doppler form a coordinate system for the surface points.

The SNR associated with the effective BIP is determined by the angle between the effective BIP and the true BIP. To lead to the following analysis, two angles of a ground target are defined here. The first is the range angle, which is the angle contained by the target, the sensor position, and the planet center. The second is the azimuth angle, which is the angle between the line of a target and the sensor and the projection of this line on the boresight plane which contains \hat{r} and \hat{c} . According to this definition the nominal range angle of the true BIP is the radar look angle and the nominal azimuth angle of the true BIP is 0. The range and azimuth angle of the effective BIP will be referred to as the effective range and azimuth angle, respectively. A pictorial definition of these angles is also shown in Figure 2. The effective pointing errors take the meaning of the differences in range and azimuth angle between the effective BIP and the true BIP.

A systematic approach to analyze the effective pointing error will be given below. The analysis contains three stages. The first stage gives the effect of the error of the ephemeris in Keplerians on the error of the ephemeris in the local coordinates.

The second stage gives the effect of the ephemeris errors in the local coordinates on the error of the BIP range and Doppler estimates. The third is the effect of the error of the BIP range and Doppler estimates on the effective range and azimuth pointing error. The analysis stages presented below will be in a reversed order since most of the reader are more interested in the radar system rather than the orbital system.

4. Effective Pointings vs Range and Doppler

To characterize the effect of errors in the range and Doppler estimates on the effective pointing errors, their partial derivatives need to be analyzed. Since the magnitude of the range and Doppler error is expected to be quite low, these partial derivatives shall be limited to the first order. This is also true for the following two stages. To formulate the partials, the nominal range and Doppler are required. As shown in Figure 3, the range of the true BIP from the sensor may be expressed by

$$R = \frac{1}{2}(b - \sqrt{b^2 - 4(H^2 - 2HR_v)}) \quad (1)$$

where b is given by

$$b = 2(H + R_v)\cos\theta_L. \quad (2)$$

The look angle θ_L is given by $\cos^{-1}(-\hat{r} \cdot \hat{p})$, where \hat{p} is the boresight pointing vector. The incidence angle may be expressed by

$$\theta_i = \theta_L + \tan^{-1}\left(\frac{R\sin\theta_L}{H + R_v - R\cos\theta_L}\right). \quad (3)$$

The Doppler frequency of the BIP is given by

$$f_d = \frac{2\vec{v}_s \cdot \hat{p}}{\lambda}. \quad (4)$$

where \vec{v}_s is the spacecraft velocity.

Denote r as the range estimate. The range pointing error $\Delta\theta_L$ associated with an error Δr in the range estimate can be solved from Figure 4. This leads to the partial derivative given below. Denote θ_r and θ_a as the effective range and azimuth angles, respectively. The range and Doppler estimates are denoted as r and f , respectively.

$$\frac{\partial\theta_r}{\partial r} = \frac{1}{R \tan\theta_i} \quad (5)$$

As long as there is no error in the range estimate, the effective range angle is equal to the true range angle. Therefore,

$$\frac{\partial\theta_r}{\partial f} = 0. \quad (6)$$

In the presence of an azimuth pointing error of $\Delta\theta_a$, the pointing vector can be expressed by $\hat{p} + \Delta\theta_a \hat{a}$. Using Equation (2), it can be seen that the Doppler change would be $2v_a \Delta\theta_a / \lambda$, where λ is the wavelength. This leads to

$$\frac{\partial \theta_a}{\partial f} = \frac{\lambda}{2v_a} \quad (7)$$

A pointing direction change confined in the boresight plane (which contains \hat{r} and \hat{c}) would change the range between the BIP and the spacecraft. It would also change the Doppler frequency if the radial component of the spacecraft velocity exists. On the other hand, an error in the range estimate in accompany with correct Doppler estimate would force the effective BIP out of the boresight plane. In such a case, a range estimate error may affect the effective azimuth angle. It can be shown that the ratio of the Doppler change and range change in the boresight plane is given by

$$\frac{df}{dr} = \frac{2v_r \sin \theta_L}{\lambda R \tan \theta_i} \quad (8)$$

Using this result and Equation (7), we have

$$\frac{\partial \theta_a}{\partial r} = \frac{-v_r \sin \theta_L}{v_a R \tan \theta_i} \quad (9)$$

5. Range-Doppler vs Local Spacecraft Status

This section gives the partials of the range and Doppler estimates with respect to the error in each component of the spacecraft position and velocity vectors. The coordinates used here is based on a local coordinate system in which the radial, along-track, and cross track directions are determined by the true spacecraft position and velocity vectors. The sensor position and velocity vectors in the local coordinate are denoted as \vec{x}_L and \vec{v}_L , respectively. Let $\vec{x}_L = [x_r, x_a, x_c]^T$ (T denotes vector transpose) and $\vec{v}_L = [v_r, v_a, v_c]^T$.

As shown in Figure 5, with the correct pointing knowledge, an error Δx_r in the spacecraft radial position would cause an error in the range estimate of Δr . From the geometry, the following result can be obtained

$$\frac{\partial r}{\partial x_r} = \lim_{\Delta x_r \rightarrow 0} \frac{\Delta r}{\Delta x_r} = \frac{1}{\cos \theta_L} + \frac{\tan \theta_L \sin(\theta_i - \theta_L)}{\cos \theta_i} \quad (10)$$

Similarly, from the geometry in Figure 6, the following result can be obtained

$$\frac{\partial r}{\partial x_c} = \frac{\sin(\theta_i - \theta_L)}{\cos \theta_i} \quad (11)$$

Based on Equation (2), the Doppler estimate can be

shown to be

$$f = \frac{2(v_r \cos \theta_L + v_c \sin \theta_L)}{\lambda} \quad (12)$$

This leads directly to the following results

$$\frac{\partial f}{\partial v_r} = \frac{-2 \cos \theta_L}{\lambda} \quad (13)$$

and

$$\frac{\partial f}{\partial v_c} = \frac{2 \sin \theta_L}{\lambda} \quad (14)$$

Since the spacecraft velocity estimate has no effect on the BIP range estimate, we can see that

$$\frac{\partial r}{\partial v_r} = \frac{\partial r}{\partial v_a} = \frac{\partial r}{\partial v_c} = 0 \quad (15)$$

Since the BIP Doppler estimate is based only on the spacecraft velocity estimate and the pointing knowledge, error in the the spacecraft position estimate will not affect on the estimate. This leads to the following result.

$$\frac{\partial f}{\partial x_r} = \frac{\partial f}{\partial x_a} = \frac{\partial f}{\partial x_c} = 0 \quad (16)$$

The pointing vector has no along-track component. This leads to $\partial f / \partial v_a = 0$. It is not difficult to see that a slight change in the along-track position does not change the BIP range estimate. Therefore, $\partial r / \partial x_a = 0$.

6. Local Spacecraft Status vs Keplerians

An elliptical orbit can be represented by six parameters referred to as the Keplerians. The first three parameters are the semi-major axis length, a , the eccentricity, e , and the true anomaly angle T . The true anomaly angle is the angle contains the object in the orbit, the orbit focal point located at the the planet mass center, and the periapsis (the apex near the planet). Obviously, the true anomaly angle is a function of time. T can be determined by a, e , and the time delay from the time of periapsis, i.e. $t - t_p$. This is through the following two equations.

$$T = 2 \tan^{-1} \left(\left(\frac{1+e}{1-e} \right)^{1/2} \tan(E/2) \right) \quad (17)$$

and

$$(t - t_p) a^{-3/2} (GM_v)^{1/2} = E - e \sin(E) \quad (18)$$

where E is the mean anomaly angle which is the angle contains the object, the orbit center, and the periapsis. G is the gravitational constant and M_v is the mass of Venus. The position and velocity in local coordinates can be determined by a, e, T , and the spacecraft speed at the periapsis, v_p .

$$x_r = \frac{a(1-e^2)}{1+e\cos(T)}, \quad x_a = 0, \quad x_c = 0 \quad (19)$$

$$v_r = v_p \frac{e \sin(T)}{1+e}, \quad v_a = v_p \frac{1+e \cos(T)}{1+e}, \quad v_c = 0 \quad (20)$$

where

$$v_p = \left(\frac{1+e}{a(1-e)} G M_e \right)^{1/2} \quad (21)$$

One often used coordinate system is the orbital coordinate system, in which, the origin is one focal point of the ellipse, one axis points to the periapsis, one axis is in the direction of the angular velocity of the orbit, and the other axis fulfils the right hand rule.

The other three parameters in the Keplerians represent the orientation of the orbit with respect to an inertial coordinate system. In Magellan mission, this system is the Venus-Mean-Equatorial 85 (VME85) coordinate system, which is based on the earth polar axis and the intersection of the mean planet equator with the mean earth equator at year 2000. The three orientation parameters are the inclination, I , the argument of periapsis ω , and the longitude node Ω . The geometry of the inertial coordinate system and the orbit is shown in Figure 7.

The status of the spacecraft in the local coordinate system X_L and in the VME85 coordinate system X_V are related by

$$X_V = \Omega I W T X_L \quad (22)$$

where

$$X_V = \begin{pmatrix} x_X & y_X \\ x_Y & y_Y \\ x_Z & y_Z \end{pmatrix}, \quad X_L = \begin{pmatrix} x_r & v_r \\ x_a & v_a \\ x_c & v_c \end{pmatrix}, \quad (23)$$

$$\Omega = \begin{pmatrix} \cos \Omega & -\sin \Omega & 0 \\ \sin \Omega & \cos \Omega & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad I = \begin{pmatrix} \cos I & -\sin I & 0 \\ \sin I & \cos I & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (24)$$

$$W = \begin{pmatrix} \cos \omega & -\sin \omega & 0 \\ \sin \omega & \cos \omega & 0 \\ 0 & 0 & 1 \end{pmatrix}, \text{ and } T = \begin{pmatrix} \cos T & -\sin T & 0 \\ \sin T & \cos T & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (25)$$

To derive the partial derivative of the ephemeris in the local coordinates with respect to the Keplerians, we must realize that a fixed local coordinate must be used. This is because that the local coordinate system changes when any one of the Keplerians changes. Lets define T, W, I , and Ω be the rotation matrices corresponding to the true Keplerians, and T', W', I' , and Ω' be the rotation matrix corresponding to the estimated Keplerians. The estimated spacecraft status X'_L in the local

coordinate defined by the true Keplerians can be determined from the estimated spacecraft status X'_L in the local coordinate defined by the estimated Keplerians through

$$X'_L = T^{-1} W^{-1} I^{-1} \Omega^{-1} \Omega' I' W' T' X'_L. \quad (26)$$

The partials of the spacecraft status elements with respect to the in-orbit Keplerians a, e , and, t_p are given by the following equation

$$\frac{\partial X'_L}{\partial p} = \frac{\partial}{\partial p} (T' X'_L) \quad (27)$$

where p denotes a, e , or, t_p . The partials of the spacecraft status elements w. r. t. the out-of-plane orbit Keplerians Ω, I, ω are given by the following equation

$$\frac{\partial X'_L}{\partial \Omega} = T^{-1} W^{-1} I^{-1} \Omega^{-1} \frac{\partial \Omega'}{\partial \Omega} I' W' T' X'_L, \quad (28)$$

$$\frac{\partial X'_L}{\partial I} = T^{-1} W^{-1} I^{-1} \Omega^{-1} \Omega' \frac{\partial I'}{\partial I} W' T' X'_L, \quad (29)$$

and

$$\frac{\partial X'_L}{\partial \omega} = T^{-1} W^{-1} I^{-1} \Omega^{-1} \Omega' I' \frac{\partial W'}{\partial \omega} T' X'_L. \quad (30)$$

The result of the partial derivatives can be found in [4].

7. Pointing Error vs Ephemeris Error

The partial derivatives of all three stages in the analysis of the effective pointing error caused by the ephemeris error in Keplerians have been given in the previous sections. Now, it is ready to formulate the pointing error estimate. The error of the Keplerians can be represented by its covariance matrix Δ .

$$\Delta = \begin{pmatrix} \sigma_a^2 & \rho_{ae} & \rho_{at_p} & \rho_{aI} & \rho_{a\Omega} & \rho_{a\omega} \\ \rho_{ea} & \sigma_e^2 & \rho_{et_p} & \rho_{eI} & \rho_{e\Omega} & \rho_{e\omega} \\ \rho_{t_p a} & \rho_{t_p e} & \sigma_{t_p}^2 & \rho_{t_p I} & \rho_{t_p \Omega} & \rho_{t_p \omega} \\ \rho_{Ia} & \rho_{Ie} & \rho_{It_p} & \sigma_I^2 & \rho_{I\Omega} & \rho_{I\omega} \\ \rho_{\Omega a} & \rho_{\Omega e} & \rho_{\Omega t_p} & \rho_{\Omega I} & \sigma_{\Omega}^2 & \rho_{\Omega \omega} \\ \rho_{\omega a} & \rho_{\omega e} & \rho_{\omega t_p} & \rho_{\omega I} & \rho_{\omega \Omega} & \sigma_{\omega}^2 \end{pmatrix} \quad (31)$$

where σ_p stands for the standard deviation of parameter p and $\rho_{p_1 p_2}$ stands for the covariance of parameter p_1 and p_2 since correlation between two parameters may generally exist.

Let the covariance matrix of the pointing error be represented by Λ , it is then given by

$$\Lambda = \begin{pmatrix} \sigma_{\theta_r}^2 & \rho_{\theta_r \theta_a} \\ \rho_{\theta_a \theta_r} & \sigma_{\theta_a}^2 \end{pmatrix} = \mathbf{A} \mathbf{B} \mathbf{C} \mathbf{\Delta} \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T. \quad (32)$$

where T denotes matrix transpose and matrices A, B, and C are defined by

$$A = \begin{pmatrix} \frac{\partial \theta_r}{\partial r} & \frac{\partial \theta_r}{\partial f} \\ \frac{\partial \theta_a}{\partial r} & \frac{\partial \theta_a}{\partial f} \end{pmatrix}, \quad (33)$$

$$B = \begin{pmatrix} \frac{\partial r}{\partial x_r} & \frac{\partial r}{\partial x_a} & \frac{\partial r}{\partial f} & \frac{\partial r}{\partial v_r} & \frac{\partial r}{\partial v_a} & \frac{\partial r}{\partial v_c} \\ \frac{\partial f}{\partial x_r} & \frac{\partial f}{\partial x_a} & \frac{\partial f}{\partial r} & \frac{\partial f}{\partial v_r} & \frac{\partial f}{\partial v_a} & \frac{\partial f}{\partial v_c} \end{pmatrix}, \quad (34)$$

and

$$C = \begin{pmatrix} \frac{\partial x_r}{\partial a} & \frac{\partial x_r}{\partial e} & \frac{\partial x_r}{\partial i_p} & \frac{\partial x_r}{\partial \Omega} & \frac{\partial x_r}{\partial \omega} & \frac{\partial x_r}{\partial \omega} \\ \frac{\partial x_a}{\partial a} & \frac{\partial x_a}{\partial e} & \frac{\partial x_a}{\partial i_p} & \frac{\partial x_a}{\partial \Omega} & \frac{\partial x_a}{\partial \omega} & \frac{\partial x_a}{\partial \omega} \\ \frac{\partial x_c}{\partial a} & \frac{\partial x_c}{\partial e} & \frac{\partial x_c}{\partial i_p} & \frac{\partial x_c}{\partial \Omega} & \frac{\partial x_c}{\partial \omega} & \frac{\partial x_c}{\partial \omega} \\ \frac{\partial v_r}{\partial a} & \frac{\partial v_r}{\partial e} & \frac{\partial v_r}{\partial i_p} & \frac{\partial v_r}{\partial \Omega} & \frac{\partial v_r}{\partial \omega} & \frac{\partial v_r}{\partial \omega} \\ \frac{\partial v_a}{\partial a} & \frac{\partial v_a}{\partial e} & \frac{\partial v_a}{\partial i_p} & \frac{\partial v_a}{\partial \Omega} & \frac{\partial v_a}{\partial \omega} & \frac{\partial v_a}{\partial \omega} \\ \frac{\partial v_c}{\partial a} & \frac{\partial v_c}{\partial e} & \frac{\partial v_c}{\partial i_p} & \frac{\partial v_c}{\partial \Omega} & \frac{\partial v_c}{\partial \omega} & \frac{\partial v_c}{\partial \omega} \end{pmatrix}. \quad (35)$$

In case that both the ephemeris and its covariance matrix are given in the local coordinate, then matrix C is not needed in evaluating the pointing covariance matrix A. Note that in estimating the variance of the effective pointing error, the off-diagonal terms in Δ cannot be ignored. Otherwise, the magnitude of the pointing error estimate tends to be worse. This can easily be seen when the ephemeris covariance matrix is given in the local coordinate system. For example, an error in one of the orbit orientation angles would cause error in all six spacecraft status components in the local coordinates. If we ignore the covariance between these components, we would get a nonzero effective pointing error estimate. However the fact is that since there is no change in the in-orbit Keplerians, the geometry remains unchanged. Therefore, the range and Doppler estimates should be correct. This results in no error in the effective pointing angles.

8. Magellan Pointing Error Estimate

The analysis results obtained in this paper are applied to the Magellan mission. The formulas of all partial derivatives were implemented in a computer program. Based on the nom-

inal Keplerian values listed below, an assumed planet radius, the mass of Venus, and a predetermined set of boresight pointing directions, the numerical values of partial derivatives can be evaluated for any selected true anomaly angle. A set of covariance matrices, characterizing the orbital estimation accuracy over a number of mission days, were supplied by the JPL Magellan navigation group. By integrating these two inputs according to Equation (35), the performance of the effective pointing angles were determined. Since significant correlation does exist among various pairs of the Keplerian parameters, this method has proven to be much more accurate than the method not considering the correlations between Keplerian parameters.

Nominal Magellan Keplerian Values are the following:

$$a=10190.3005326 \text{ km}, e=0.3816669116067,$$

$$I=86^\circ, \Omega=68.9^\circ, \text{ and } \omega=170^\circ$$

The semi-major axis length a and the eccentricity e are obtained based on the design of a 3.15 hour orbit period and a periapsis altitude of 250 km.

REFERENCE:

- [1] Dallas, S. S. and Nickle, N. L., The Magellan Mission to Venus, "Advances in Astronautical Sciences" Part 1 titled "Aerospace Century 21", AAS 86-331, Volume 64, 1987.
- [2] Johnson W. and Edgerton, A., Venus Radar Mapper (VRM) Multimode Radar System Design, SPIE Vol. 589 Instrumentation for Optical Remote Sensing from Space (1985)
- [3] M. Chen and M. Jin, Magellan SAR Processing Algorithm and II/W Design, IGARSS Symposium 1989.
- [4] M. Jin, Effective Pointing Errors Induced by Ephemeris Error for a Spaceborne SAR, JPL publication, 1989.

* The research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

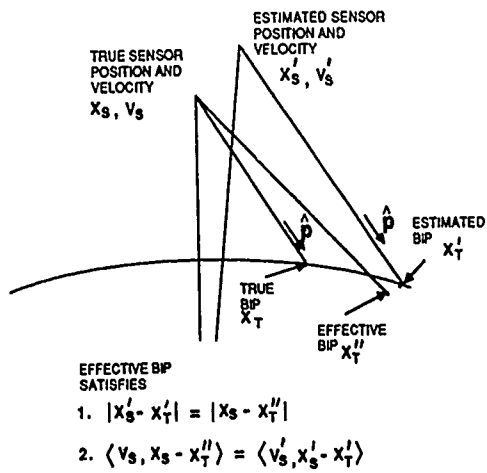


Figure 1

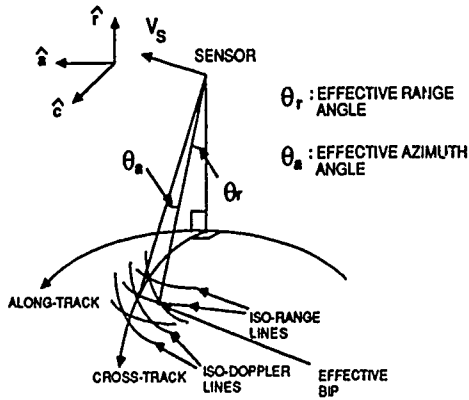


Figure 2

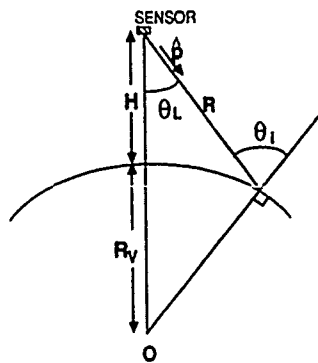


Figure 3

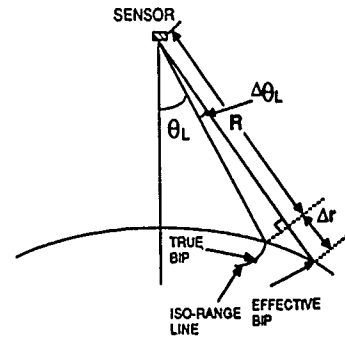


Figure 4

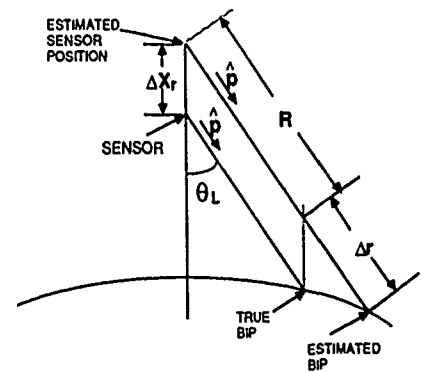


Figure 5

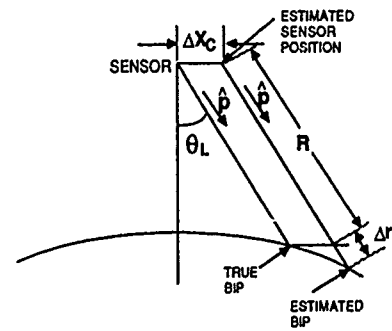


Figure 6

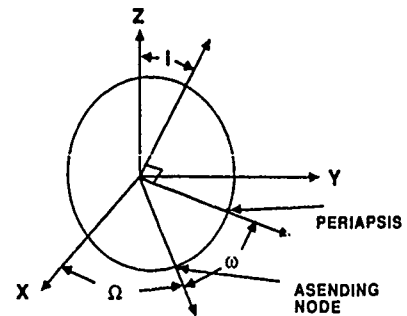


Figure 7

STATISTICALLY BASED UNSUPERVISED HIERARCHICAL IMAGE SEGMENTATION ALGORITHM WITH A BLURRING CORRECTOR

Lee, S. and Crawford, M. M.

Department of Mechanical Engineering
and
UT Center for Space Research
The University of Texas at Austin, TX, USA

Abstract.

A multistage algorithm which makes use of spatial contextual information in a hierarchical clustering procedure has been developed for unsupervised image segmentation. A Markov random field model is employed to enforce local spatial smoothness, while the maximum entropy principle is utilized to quantify global smoothness in the image. A multi-window approach implemented in a pyramid-like data structure which uses a so-called boundary blocking operation is utilized to increase computational efficiency. Schwarz information criterion is suggested as a means of selecting the level in the clustering hierarchy which corresponds to the optimal state. A Bayesian-based blurring corrector has been developed to iteratively recover images from blurred versions of actual pixel values.

Keywords: unsupervised image segmentation; hierarchical clustering; Bayesian estimation; Markov random field; maximum entropy principle; pyramid structure.

1. Introduction

A variety of techniques for image segmentation have been developed for many scientific applications including remote sensing. The most accurate segmentation approaches utilize the maximum likelihood or the Bayesian performance function which both require *a priori* knowledge for parameterized models of classes in images. The Bayesian method presented in this paper is based on unsupervised learning of the values of parameters which must be known in order to carry out the segmentation.

Hierarchical clustering (Anderberg, 1973) is the most plausible and well-grounded tool for the analysis of signature-dependent noise models as well as for unsupervised analysis. The conventional "distribution-free" or "context-free" similarity measures between regions in typical clustering approaches fail to include important information about image spatial structures that should be exploited when the scene is segmented. The original radiated intensity of a pixel is a random variable whose measurement is contaminated by energy radiated from nearby pixels. This results in spatial blurring which causes the whole observation process to be spatially autocorrelated. For most scenes, pixels that are spatially close tend to have the same or similar intensities. Because of this spatial smoothness, contextual information can be used to advantage in allocating pixels to a class. Most of statistical image segmentation methods which incorporate this idea

use Markov random fields, equivalently *Gibbs* random fields (GRF) (Derin and Elliott, 1987) to characterize geophysical connectedness. While the GRF represents local characteristics of the image structure, *entropy* (Jaynes, 1968) can be interpreted as an alternative measure which embodies the spatial smoothness in a global sense. Both are incorporated in a new multistage algorithm which has been developed.

Analysis of multispectral image data from a large area is computationally intensive with a hierarchical clustering approach. A suboptimal algorithm has been developed to increase the computational efficiency of the regional-clustering segmentor. It uses a pyramid-like structure in a way that exploits regional contiguity in merging clusters. A process called boundary blocking is applied to individual nonoverlapping windows of a large image which are successive subdivisions in the hierarchical representation of the image.

Conventional approaches for determining the number of classes in an image use the likelihood ratio and "Akaike's information" (Solve, 1983). These approaches are insensitive to image size and fail to find parsimonious models. Alternatively, Schwarz information criterion (SIC) is suggested as a reasonable means for determining the optimal state. The components of the algorithm are outlined in the following sections. The results of the classification are reported for a simulated process.

2. Algorithm Outline.

The components of the algorithm are outlined in Figure 1. The first component is called the regional-clustering segmentor. It applies hierarchical clustering using the *maximum a posteriori* (MAP) criterion with a restriction that pixels in a cluster should be contiguous. The global-clustering segmentor in the next stage merges clusters according to the maximum entropy principle (Frieden, 1972) among all possible combinations. Both segmentors contain a smoothing-parameter estimator and Bayesian-based blurring corrector. The coefficients, which represent the optimal strength of prior beliefs with respect to the information provided by the realization, are estimated based on an initial configuration. The Bayesian-based blurring corrector is designed to iteratively enhance the observed image using adaptive estimates of the parameters. The segmentor system then partitions the image on the basis of these enhanced intensities rather than of the observation.

3. Digital Image Model.

A sample image is considered as a set of n pixels, where each pixel is represented by an intensity vector of d spectral components. The image process is assumed to be of the form

$$Y = B \cdot X(\omega) + \varepsilon \quad (1)$$

where ω is an integer valued random vector associated with a particular configuration of classes in an image, X is a mapping operator of ω into real intensity values, B is a blurring operator, and ε is an independent noise field. In modelling the process of (1), simplifying assumptions are used to provide numerous advantages in mathematical manipulations while still retaining the essence of the physical process being represented: (a) the original intensity is independently distributed according to a composite multivariate Gaussian which is identical within a region class, (b) the original distribution of object radiated intensity is acted upon by a homogeneous linear system to produce a blurred image and (c) the noise process is generated by different independent and identical Gaussian field in each spectral band.

Let $I_n = \{1, \dots, n\}$ be a set of indices of pixels, and let $J_m = \{1, \dots, m\}$ be a set of indices representing the classes whose associated partition of I_n is $G = \{G_j, j \in J_m\}$. If the intensity of pixels is distributed $X_i \sim N(\mu_j, \Sigma_j)$ for $i \in G_j \in G$, the conditional distribution of X , given $\omega = \{\omega_i : i \in I_n \text{ and } \omega_i \in J_m\}$,

$$f(x|\omega) = \kappa \exp\{-\frac{1}{2} \sum_{j \in J_m} Q_x(G_j)\}$$

$$Q_x(G_j) = \sum_{i \in G_j} \log |\Sigma_j| + (x_i - \mu_j)' \Sigma_j^{-1} (x_i - \mu_j) \quad (2)$$

where κ is a normalizing constant. Suppose that multispectral images are blurred according to assumption (b), and the boundaries of sample images are defined on a "periodic random field". The blurring operator $B = \text{circulant}\{B_{ij}\}$ can then be expressed as a block circulant and stochastic matrix where $B_{ij} = \text{diagonal}\{b_{ijk}, k = 1, \dots, d\}$, $\{b_{ijk}, j = 1, \dots, n\}$ corresponds to the point spread function of the i th pixel in the k th spectral band and $\sum_{j=1}^n b_{ijk} \geq 0 = 1, \forall i, k$. If $\{\varepsilon \sim N(0, \Sigma_\varepsilon)\}$ where Σ_ε is diagonal, the distribution of the observation Y , conditioned on X , is given by

$$f(y|x) = \kappa \exp\{-\frac{n}{2} \log |\Sigma_\varepsilon| - \sum_{i=1}^n Q_{y_i}(x)\}$$

$$Q_{y_i}(x) = (y_i - \sum_{j=1}^n B_{ij}x_j)' \Sigma_\varepsilon^{-1} (y_i - \sum_{j=1}^n B_{ij}x_j). \quad (3)$$

4. Bayesian Objective Function

The image process is supposed to combine a discrete random field for the region class process and a continuous random field for the intensity process and the noise process. An approach which integrates multiple processes by Bayes' theorem is applied to estimate the classes of pixels for an image. Given the observed process Y , finding the mode of the posterior probability distribution (MAP) of ω is equivalent to maximize the log penalty function

$$\ell PN = \log P(Y|\omega) + \log P(\omega). \quad (4)$$

The second term is known as "roughness penalty" or smoothing term. In the remainder of this section, objective measures are discussed for determining the optimal segmentation of "multiple compound stochastic" image processes based on the Bayesian approach. The GRF and the maximum entropy principle are used to quantify the spatial smoothness probabilistically, that is, to provide some type of prior information of the class process.

Let $R = \{R_i, i \in I_n\}$ be a neighborhood system for I_n .

For a probability structure of ω , a class of Gibbs measures with respect to $\{I_n, R\}$ has been introduced with energy function $U(\omega, \mu)$: if $|\cdot|^2$ denotes the square of each element,

$$P(\omega) = \kappa_0 \exp\{-U(\omega, \mu)\} \\ U(\omega, \mu) = \sum_{(r,s) \in C_p} \alpha'_{rs} |\mu(\omega_r) - \mu(\omega_s)|^2 \quad (5)$$

where μ is a vector function of the mean intensity vector associated with the class, α_{rs} is a nonnegative coefficient vector, and κ_0 is the normalizing constant. C_p denotes the collection of all "pair cliques" relative to $\{I_n, R\}$ which are subsets of I_n such that a clique involves a pair of indices, and two pixels corresponding to the indices are neighbors. In the energy function, the parameter α_{rs} represents the "bonding strength" of the r th pixel and s th pixel. Using this Gibbs measure as a probability model of ω in (4), the spatial smoothness is incorporated into the Bayesian approach by penalizing the objective function as the difference in spatial proximity of the mean intensity increases.

Suppose that a mean intensity associated with a class is the "average radiant energy" of a unit cell in the corresponding region, resulting from the random radiation of the particles. Under this model, the entropy for a single spectral band can be defined as

$$H(P) = -\phi(\omega, \mu) = - \sum_{i \in I_n} \frac{\mu(\omega_i)}{T_\mu} \log \frac{\mu(\omega_i)}{T_\mu}, \quad (6)$$

where $T_\mu = \sum_{i \in I_n} \mu(\omega_i)$. Given a particle from the observed image space, H represents the uncertainty as to which pixel radiated it. Maximizing H has a smoothing influence on the mean intensity level of the image (Orlando, 1987).

If particles act independently in different spectral bands, the maximization of entropy for individual bands then results in an independent smoothing influence on the corresponding spectral intensity levels with a independent manner. Since these multiple objectives potentially result in conflicts, the final estimation should accommodate trade-offs between objectives. Under a constraint of consistency with the observed data, a maximum entropy approach for the segmentation is considered as a constrained optimization problem with the multiobjective function $f(\phi) = \alpha' \phi$ where α is a nonnegative weighting vector and ϕ is a vector which contains the values of ϕ 's in (6) of the corresponding spectral band. It is equivalent to the MAP estimation with $P(Y|X) \propto \exp\{-\Delta_Q\}$ and $P(\omega) \propto \exp\{-\alpha' \phi\}$, where Δ_Q is a measure between μ and the observed image.

Assuming that Δ_ω (Gibbs measure or entropy) contains proper information about the behavior of the region class process, the MAP estimation can be restated as an optimization problem:

$$\min \{\Delta_Q + \beta \Delta_\omega\}. \quad (7)$$

Using a Bayesian interpretation β is referred as a smoothing parameter which represents the relative strength of prior beliefs compared to the information in the observed data. The smoothing parameter then indicates the desired trade-off between two counteractive objectives.

5. Hierarchical Clustering.

The regional segmentor is designed to merge adjacent re-

gions while the global segmentor classifies regions into same class regardless of their geographical positions. The hierarchical clustering procedure is applied to both segmentors using adaptive parameters computed from the updated regional configurations at each step.

Let $G_h = \{G_j^h, j = 1, \dots, m_h\}$ be a partition at the h th step, and let the corresponding configuration of the pixel-classes be ω_h . If $\omega_h^{r,s}$ is the configuration resulting from merging the r th and s th regions, then the similarity coefficient for G_r and G_s can be derived from the the objective function of (7):

$$\begin{aligned} \nu_{r,s}^h &= \delta_Q^h(r, s) + \beta \delta_\omega^h(r, s) \\ &= [\Delta_Q(y|\omega_h^{r,s}) - \Delta_Q(y|\omega_h)] + \beta [\Delta_\omega(\omega_h^{r,s}) - \Delta_\omega(\omega_h)]. \end{aligned}$$

It is proportional to the conditional posterior probability

$$P(\omega_h^{r,s} | G_h \setminus G_r^h \cup G_s^h, Y).$$

The merge decisions with this criterion yield MAP estimates among possible states at each step in the clustering approach.

It is desirable that the similarity coefficients be computed by locally updating them to correspond to the configuration resulting from every iteration. This requirement is very natural for δ_ω when the entropy measure of (6) applies to Δ_ω . Using the Gibbs measure of (5) for Δ_ω , the δ_ω of the coefficients in the spatial proximity of the new boundary formed by merging are updated. However, if Δ_Q is a direct distance measure between the mean intensity and observation in (1), the local updating condition is not satisfied because of the spatial component of the whole observation process. In order to overcome this problem, the Bayesian-based blurring corrector produces a deburred image \hat{x} which is assumed to have no spatial component. If \hat{x} is an appropriate estimate of x , the MAP clustering classifier can employ similarity coefficients based on \hat{x} rather than y . With the Gaussian assumption, δ_Q is simply

$$\delta_Q^h(r, s) = Q_x(G_r^h \cup G_s^h) - Q_x(G_r^h) - Q_x(G_s^h)$$

The parameters related to the merge can be estimated by maximizing the likelihood conditioned $\omega_h^{r,s}$. The $\delta_\omega^h(r, s)$ for the regional segmentor is equivalent to the energy change at the boundaries and for the global segmentor to the increase in entropy portion related to G_r^h and G_s^h associated with changes in the estimate of the mean intensity resulting from merging two regions.

In image segmentation, the configuration ω should be estimated such that the corresponding mean intensity μ results in fitting to the observed data to some degree. This requirement can be satisfied by imposing some constraint on the distance Δ_Q in (7). Then, the smoothing parameter can be chosen by using the optimization approach (Lee, 1989).

When the cluster analysis is used with parametric models, the choice of number of classes is usually based on generalized likelihood ratio test using the asymptotic chi-squared distribution. But, if a more parsimonious model is true, the distribution of the test statistic may not, even asymptotically, be chi-squared. Akaike's information criterion (Schlove, 1983) has also been usually used in image analysis. This statistic has a similar problem to the chi square distribution of the likelihood ratio (Titterton, 1981). Schwarz (1978) suggested an alternative to Akaike's approach for the model selection criterion:

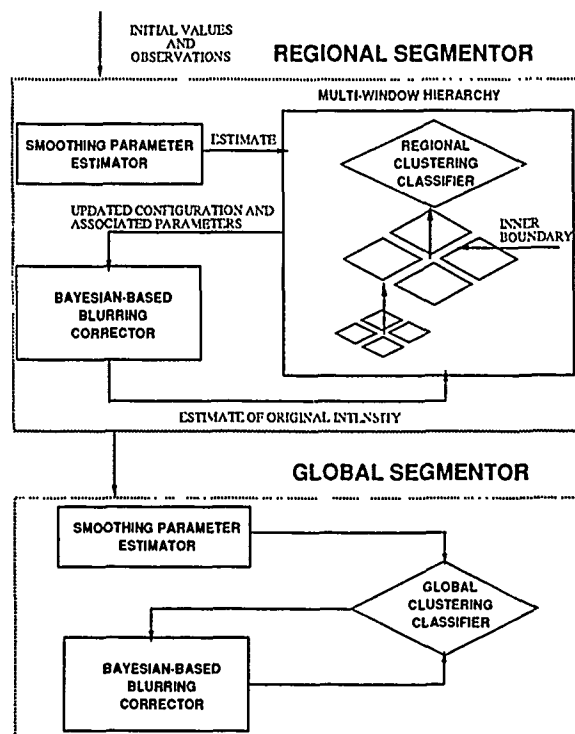


Figure 1. The multistage clustering image segmentor.

$$SIC = -2 \log(\text{maximum likelihood}) + m_p \log n$$

where m_p is the number of independent parameters estimated when using m classes and n is the number of independent observations. The SIC criterion is a proper approach for choosing number of classes in images because they usually involve large data sets and are envisioned to have a relatively smaller number of classes as they become large. The clustering approach incorporates this criterion for the threshold to prevent the pathological merging of regions which are unacceptably different in corresponding observations, and eventually for the decision of stopping the clustering.

6. MAP Restoration.

If the blurring is governed by B which is a stochastic matrix, and the distribution of the original intensity is a composite Gaussian, the conditional distribution is

$$f(x|y) \propto (y - Bx)' \Gamma_y^{-1} (y - Bx) + (x - \mu)' \Gamma_x^{-1} (x - \mu)$$

where $\Gamma_y = \text{diagonal}\{\Sigma_\epsilon\}$ and $\Gamma_x = \text{diagonal}\{\Sigma_i\}$ are $d \times n$ covariance matrices corresponding to the measurement noise and the original intensity. Then, the MAP estimate of X is given by:

$$\hat{x} = (I + \Psi_v B' B)^{-1} (\Psi_v B y + \mu) \quad (11)$$

where $\Psi_v = \Gamma_x \Gamma_\epsilon^{-1}$. If the spectral radius of Ψ_v is sufficiently smaller than 1, (11) can be approximated by (Pullman, 1976):

$$\hat{x} \approx \mu + \Psi_v B y - \Psi_v B' B \mu + o(\Psi_v).$$

This would be a good estimate of the original intensity for images where the variation of the intensity representing physical properties of region classes is generally much smaller than the variance of the noise produced during collecting and transmit-

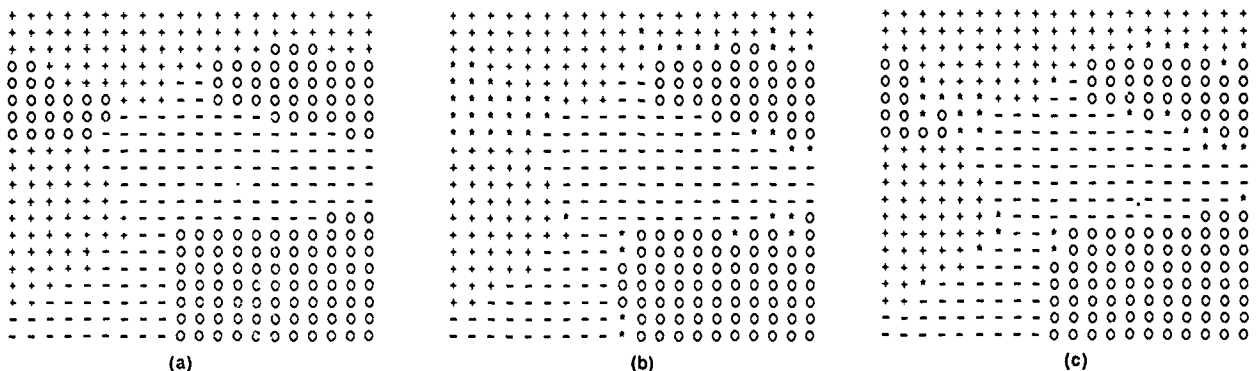


Figure 2. Results of the classification for a two dimensional 20 by 20 simulated image (* s are misclassified pixels).

ting of data by atmospheric effects.

Based on this approximation, the Bayesian-based blurring corrector has been designed for the clustering classifier. Details of the corrector will be described in a forth coming paper.

7. Multistage Regional Clustering.

The new hierarchical clustering algorithm is acclimated to the merges restricted to contiguous regions and the similarity coefficients satisfying the local updating condition. While the best merge is sought among all possibilities in the conventional approach, the new approach looks for the closest pair in the storage vectors of a given dimension s which contain the information about pairs corresponding to the s largest similarity coefficients. This information is updated at each iteration, using the local updating rule. The searching efficiency relative to the best case of the conventional approach is approximately $s/2m_0(m_0 + 1)$ where m_0 is the initial number of clusters. Nevertheless, the algorithm's "expected time complexity" is exponentially proportional to the initial number of clusters due to updating the regional information (for example, neighboring clusters and boundary) at each iteration. The multistage strategy can be implemented by constructing a pyramid-like hierarchy system. This could be far from the optimal solution if the clustering were independently applied to the nonoverlapping windows generated from the hierarchy system. To overcome this problem, a boundary blocking operation has been developed by defining "contagious clusters" in the "inner boundary" (boundary between windows as shown in Figure 1) at the initial iteration. It is based on the idea that regions located in different window should be combined through a common separating boundary. If a cluster "touches" a contagious cluster (that is, the merge of these clusters is considered as the best at that iteration), it also becomes contagious, and the operation eliminates the pairs which are both contagious from a set of pairs feasible to merge.

8. Examples and Conclusions.

The new multistage algorithm including the Bayesian-based blurring corrector has been tested with simulation data and agricultural SPOT imagery of the fields in South Texas. Figure 2 illustrates examples of the results for a simple simulation image of two spectral bands. The data generated from a Markov random field of 3 classes (the original pattern is shown in Figure 2-a) were blurred, and independent

noise was added to the blurred version. The signal-noise ratio was approximately 1.3 in the simulated data. First, the data were classified with a version of the clustering algorithm whose similarity measure doesn't include contextual information. Figure 2-b shows the classified image. Figure 2-c shows the results of the classification using a similarity coefficient incorporating spatial smoothing. The error rates are 11% and 7.5% respectively.

These results reinforce the importance of the utilization of contextual information in image segmentation. The blurring corrector provides more reliable information about the intensity process which in turn results in a more precise classification. In addition to making the hierarchical clustering algorithm useful for the analysis of large remotely sensed images, the multistage operation often leads to better results than those of the single stage operation because unwholesomely large clusters are prevented from growing in early iterations.

Reference.

- Anderberg, M. R., *Cluster Analysis for Application*. Academic Press, New York, 1973.
- Derin, H. and Elliott, H., "Modelling and segmentation of noisy and relaxation algorithms for segmenting non-causal Markovian random fields," *IEEE Trans. Pattern Anal. Machine Intell.* Vol. PAMI-9, 39-55, 1987.
- Frieden, B., R., "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.* Vol. 62, 512-518, 1972.
- Jaynes, E. T., "Prior probabilities," *IEEE Trans. System Sci. and Cyber.* Vol. SSC-4, 227-241, 1968.
- Lee, S., "Statistically based unsupervised classification and adaptive estimation in multispectral, multitemporal image processing," *Ph. D. Dissertation*. Univ. of Texas at Austin, 1989.
- Pullman, N. J., *Matrix Theory and Its Applications*. Marcel Dekker Inc., New York, 1976.
- Schwarz, G., "Estimation of the dimension of a model," *Anal. Math. Statistics*. Vol. 6, 461-464, 1978.
- Scolve, S. L., "Application of the conditional population-mixture model to image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.* Vol. PAMI-6, 428-432, 1983.
- Titterton, D. M., Comments on "Application of the conditional population-mixture model to image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.* Vol. PAMI-6, 656-657, 1984.

PIXEL RELAXATION LABELLING USING A DIMINISHING NEIGHBOURHOOD EFFECT

T. LEE

Department of Electronics
The Chinese University of Hong Kong
Shatin NT
Hong Kong

J.A. RICHARDS

Department of Electrical Engineering
University College
University of New South Wales
Campbell ACT 2600
Australia

ABSTRACT

When applied to pixel data, particularly in remote sensing, normal schemes for probabilistic relaxation labelling lead to an improvement in labelling accuracy with iteration followed by a degradation, sometimes leading to a final result worse than that initially. A simple modification of the relaxation process is described that allows spatial consistency to be developed while avoiding labelling degradation. Based upon decreasing the neighbourhood influence with iteration count, it permits an assessment *a priori* of the optimum iteration at which the process should be terminated. Results are presented based on modelled data, and Landsat imagery, that demonstrate the value of the technique.

KEYWORDS

Classification, Relaxation_labelling, Neighbour_weights.

1. INTRODUCTION

Spatial information is a particularly important source of ancillary data that can be exploited to enhance the accuracy of multispectral classifications, since it can describe how compatible the labels assigned to pixels in a given neighbourhood happen to be. This has particular significance in regions containing cultural or cultivated features in which the spatial properties are quite explicit.

An iterative method that has been investigated widely for developing spatial consistency in a spectrally derived thematic map is probabilistic relaxation labelling [1,2]. This requires each pixel in an image to be described by a probability vector which indicates the relative likelihoods of the various possible labels being the correct one for the pixel. Normally this vector is determined at the penultimate stage of maximum likelihood estimation, just prior to the maximum selection step. The label probability vector for each pixel is then updated iteratively in the relaxation process by incorporating information from near neighbouring pixels. The link to the neighbours is via a set of compatibility coefficients, that describe the likely mutual labellings on a pair of pixels, and a set of neighbour weights that express the relative importances of the neighbours. The compatibility coefficients are derived either from a known spatial model for the region in question or are estimated from training data.

Labelling accuracy in a relaxation process has been observed to improve during early iterations and then degrade [3], sometimes leading to a poorer result than the initial labelling if the iterations are not terminated. Unfortunately, in a practical exercise, the point of termination is not known. Consideration must be given

therefore to controlling the degradation phase if the technique is to be useful as a means for exploiting spatial information.

Several approaches have been adopted in attempting to circumvent the degradation process by controlling the relaxation parameters [3], by using non-stationary compatibility coefficients [4], or by using only a single iteration, the last of which is based upon the observation that most (but not all) improvement occurs at this stage. These suffer through being non-optimal or by their excessive computational demands.

Here a method to overcome label degradation is proposed in which the influence of neighbouring pixels is reduced with the iteration count. This is motivated by a consideration of how information propagates with iteration count [5] and the extent of label correlation known to exist for particular data sources [6]. The diminishing neighbourhood effect is implemented by a simple scheme for modifying the neighbour weights for each iteration. This is readily implemented and has a theoretical guarantee that the relaxation process will be bounded by the standard and the noniterating processes.

II. DIMINISHING NEIGHBOURHOOD EFFECT

The updating rule for the standard relaxation algorithm is

$$p_i^{k+1}(\lambda) = \frac{p_i^k(\lambda) Q_i^k(\lambda)}{\sum_x p_i^k(\lambda') Q_i^k(\lambda')} \quad (1)$$

with a neighbourhood function $Q_i^k(\lambda)$ defined as

$$Q_i^k(\lambda) = \sum_{j \in J} d_{ij} \sum_x p_j(\lambda | \lambda') p_j^k(\lambda') \quad (2)$$

in which p_i is the label probability vector for the i th pixel, J denotes the set of neighbouring pixels about and including the centre pixel i and $p_{ij}(\lambda | \lambda')$ denotes the measure of compatibility of pixel i being labelled λ and pixel j being labelled λ' . The $p_{ij}(\lambda | \lambda')$ s are appropriately called the compatibility coefficients. Peleg [7] has recommended a different set of compatibility coefficients based upon label probabilities conditional upon neighbourhood measurement vectors. The neighbour weights, which are denoted by d_{ij} in (2), represent the relative amount of influence in the updating process coming from the corresponding member of the neighbourhood.

Based upon Haralick's [5] interpretation that the relaxation process expands the effective contributing context with iterations,

and that the process starts to degrade at 4 or 5 iterations, there appear to be two ways to minimise the degradation effect at high iterations. The simplest approach would be to stop the process at iteration 4 or 5. A better approach is to decrease the neighbourhood influence with iterations, by decreasing the neighbourhood weights for neighbouring pixels according to iteration count such that these converge to zero, while the neighbour weight for the centre pixel should converge to unity.

III. A MODIFICATION SCHEME

Suppose we choose the neighbour weights, d_{ij} , in the standard relaxation algorithm [1] to be decreasing functions of iteration count, k , in the manner

$$d_{ij}(k) = d_{ij}(1) \exp^{-\alpha(k-1)}$$

so that

$$d_{ii}(k) = 1 - \sum_{j \neq i} d_{ij}(k)$$

in which $d_{ij}(1) = (1 - d_{ii}(1))/J$ and J is the number of elements in J

Here $d_{ij}(k)$ denotes the value of d_{ij} at iteration k . This choice for $d_{ij}(k)$ has the following desirable properties:

1. as $k \rightarrow \infty$; $d_{ij} \rightarrow 1$, $d_{ij} \rightarrow 0$
2. $d_{ij}(k+1) \leq d_{ij}(k)$ for all $k \geq 1$
3. standard relaxation is a special case given by $\alpha = 0$
4. non-iterating relaxation is a limiting case for $\alpha \rightarrow \infty$

Equal weights for neighbouring pixels have been specified here because this is generally the case and has been implemented in our experiments.

The proposed neighbour weight functions depend on a rate parameter, α , that determines how quickly the neighbourhood effects are diminished. With α small, the algorithm is allowed to iterate relatively freely and at the limit, standard relaxation results. With α large, the converse applies and at the limit, the algorithm becomes noniterating. Consequently, relaxation with the modified scheme proposed will have performance bounded by those of the standard algorithm and the noniterating version.

Although there is no explicit stopping criterion for relaxation with the proposed modification scheme, we have chosen to terminate the process when the d_{ij} 's become negligible ($<10^{-7}$) as the process iterates. This allows the maximum number of iterations required for each combination of α and $d_{ij}(1)$ to be predicted.

IV RESULTS

Experiments on the diminishing neighbourhood relaxation rule were carried out with a model data set and also with a Landsat MSS image segment of 82×100 pixels in an agricultural region. The model data set consists of only two classes, hypothetically named as 'w' and 'b'; the ground truth map is shown in Fig. 1a. Initial probability estimates for the relaxation process and the compatibility coefficients were generated from this map. The initial probabilities were determined arbitrarily from a random Gaussian generator, assuming distributions that have mean = 0.7 and standard deviation = 0.1 for label 'w', and for label 'b', a mean = 0.6 and a standard deviation = 0.1.

The experiment with the Landsat MSS image provided a test of the algorithm for multiclass labelling with the initial probability

estimates derived from spectral information and ground cover statistics, while the experiments on model data give better insight into the effect of the algorithm on field geometry.

In all experiments, $d_{ij}(1)$ was chosen to be 0 and Peleg's coefficients [7] were used. The initial labelling for the model data set is depicted in Fig. 1b, showing the initial classification error to be 21 out of the map of 100 pixels. The error was reduced to 7 after one iteration of the relaxation process. The labelling after 1 iteration is shown in Fig. 1c. The full set of experimental results for different values of α are tabulated in Table 1. Wherein it is seen that the optimal value for α is 0.5; the corresponding labelling is shown in Fig. 1d.

For the Landsat MSS image in Fig. 2a, a careful photointerpretation was carried out using the four multispectral bands. This was used as a reference map for the experiment. The image was segmented into seven classes as in Fig. 2b. Means and covariance matrices for normal class density distributions were computed by extracting training areas corresponding to each ground cover type from the reference map. The normal probability distribution functions were then used to compute the initial probability estimates for the relaxation process. The labelling based on those initial estimates had an accuracy of 65.6% which was improved to 72.2% by the controlled relaxation process. The reference map was also used to tabulate the joint statistics from which Peleg's coefficients were computed.

The experimental results are summarized in Table 2 and the final labelling for the optimal α of 2.0 is depicted in Fig. 2c.

V. CONCLUDING REMARKS

From the experiments performed with the modification scheme proposed, a proper choice of α will achieve and sustain a minimum error level. Although the method of determining the optimum value of α is yet to be derived, the result from the Landsat image suggests that a range of α will give an improvement to the algorithm.

REFERENCES

1. Rosenfeld, A., Hummel, R.A., & Zucker, S.W., "Scene Labelling by Relaxation Operations", *IEEE Trans. Systems, Man, and Cybernetics*, Vol. SMC-6, pp420-433, 1976.
2. Yamamoto, H., "Some experiments on LANDSAT Pixel Classification using Relaxation Operators", *Computer Graphics and Image Processing*, Vol. 13, pp31-45, 1980.
3. Richards, J.A., Landgrebe, D.A., and Swain, P.H., "On the Accuracy of Pixel Relaxation Labelling", *IEEE Trans. Systems, Man, and Cybernetics*, Vol. SMC-11, pp303-309, 1981.
4. Kalayeh, H.M. and Landgrebe, D.A., "Adaptive Relaxation Labelling", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, pp369-372, 1984.
5. Haralick, R.M., "An Interpretation for Probabilistic Relaxation", *Computer Graphics and Image Processing*, Vol. 22, pp388-395, 1983.
6. Kettig, R.L., "Computer Classification of Remotely Sensed Homogenous Objects", LARS Information Note #050975, Purdue University, West Lafayette, IN, May 1975.
7. Peleg, S and Rosenfeld, A., "A New Probabilistic Relaxation Procedure", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, pp362-369, 1980.

TABLE 1
Error at each iteration of the algorithm with different values
of α for the model data set

| ITERATION COUNT | | | | | | | | | | | | | | |
|-----------------|---|---|---|---|----|----|----|----|----|----|----|----|----|--|
| α | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 13 | 14 | 15 | 30 | |
| 0.0 | 3 | 6 | 6 | 7 | 10 | 10 | 12 | 12 | 12 | 12 | 12 | 12 | 17 | |
| 0.1 | 3 | 5 | 5 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | |
| 0.2 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |
| 0.3 | 5 | 4 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | |
| 0.4 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |
| 0.5 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| 0.6 | 5 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| 0.7 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | |
| 1.0 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |
| 2.0 | 7 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | |

TABLE 2
Summary of the experimental results for the Landsat MSS data

| α | OPTIMAL RESULT | | | FINAL RESULT | | |
|----------|----------------|------|--|--------------|------|--|
| | % acc. | iter | | % acc. | iter | |
| 0.0 | 72.2 | 4 | | 70.6 | 32 | |
| 1.0 | 72.2 | 4 | | 71.4 | 17 | |
| 1.8 | 72.2 | 4 | | 72.1 | 10 | |
| 2.0 | 72.2 | 4 | | 72.2 | 9 | |
| 2.2 | 72.2 | 4 | | 72.2 | 8 | |
| 2.5 | 72.2 | 5 | | 72.2 | 7 | |
| 3.0 | 72.2 | 4 | | 72.2 | 6 | |

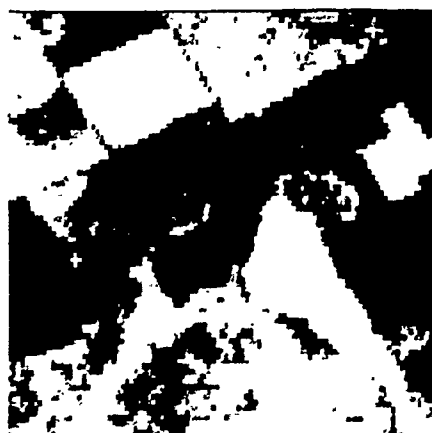


Fig. 2 (a) Band 7 of the Landsat MSS image for experiments

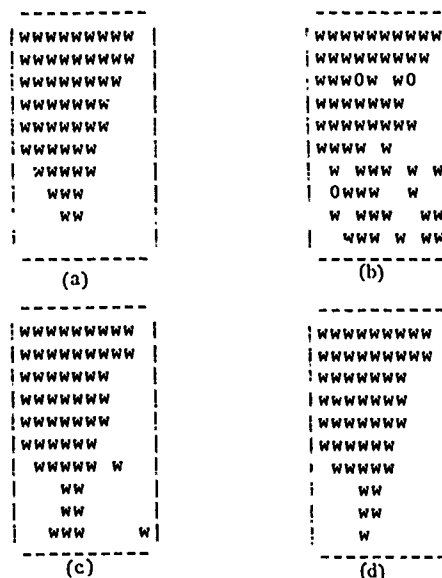


Fig. 1 (a) Ground truth for model data.
(b) Initial labelling for model data; pixels which have equal
initial probability estimates for both classes are
denoted as "0".
(c) Labelling after 1 iteration for model data.
(d) Optimal labelling for model data $\alpha = 0.5$ at iteration 3.



Fig. 2 (b) The hand segmentation of the study area. The classes
are (.) red soil (*) cotton (l) bare soil (low moisture
content) (l) bare soil (dry) (+) early vegetation growth
(X) mixed bare soil (some red soil overlying) (-) bare
soil (moist or ploughed).

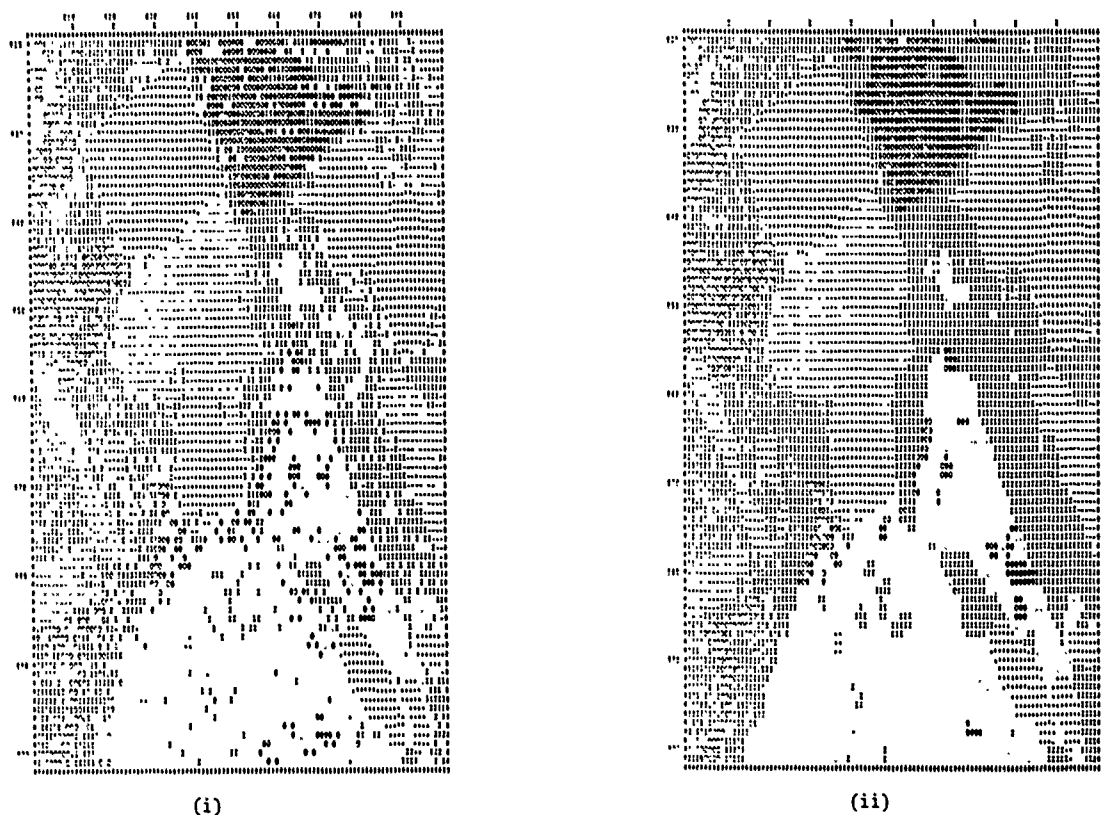


Fig. 2 (c) (i) initial labelling (ii) final labelling for $\alpha = 2.0$

SPECTRAL MATCHING IN SUPERVISED CLASSIFICATION OF TERRAIN FEATURES

Zhenkui Ma and Charles E. Olson, Jr.

School of Natural Resources
The University of Michigan
Ann Arbor, MI 48109-1115, USA

ABSTRACT

A classification algorithm has been developed which permits use of spectral patterns in the classification process. Based on a correlation decision rule, the classifier assigns an unknown pixel to the classification category having the most similar spectral pattern -- the "Nearest Spectrum." The Nearest Spectrum classifier can be used as either a parametric or non-parametric classification algorithm with respect to the training information used. It is free of statistical assumptions and sampling limitations due to the nature of the classifier.

Keywords: Classifier, Spectral Matching, Correlation

INTRODUCTION

Supervised classification of Thematic Mapper data uses statistical parameters calculated from samples of each target class to assign pixels to the most similar class. The advantages of supervised classification are quick outputs, an immediately interpretable classified map, and target oriented data processing. Disadvantages of supervised classification include the need for sufficiently large data sets to permit unbiased estimation of population parameters, and difficulties in meeting the requirements of data distribution and statistical assumptions (James, 1985). Many supervised classifiers are used in remote sensing data classification while ignoring these restrictions.

Assignment of pixels to clusters, or cover type categories, has been the primary focus of land cover/use specialists (Chuvieco and Congalton, 1988), and strategies have been developed to improve existing methods (Wharton and Turner, 1981). Using supervised or unsupervised classification algorithms for classification of remote sensor data requires a decision rule for grouping data into clusters. Bayes' rule is widely applied for probability related supervised classification algorithms, to minimize the probability of misclassification (Swain and Davis, 1978). Euclidian distance is used in some supervised classification algorithms and also in some unsupervised clustering algorithms. Many types of classification algorithms have been derived applying Bayes' rule or Euclidean distance (Skidmore and Turner, 1988), but none satisfy all needs. Our search for a better classifier included comparisons of several available classification procedures (Justice and Townshend, 1982; Tom and Miller, 1984), and a search for better sam-

pling designs for obtaining statistics to train a supervised classifier (Chuvieco and Congalton, 1988.) We also developed a data set to evaluate accuracy of a classified map (Hay, 1979), for when a training set does not truly represent a population, classification performance may be seriously limited (Hixson et al., 1979). A classification rule with as few limitations as possible was desired for developing a new classification algorithm.

THE NEAREST SPECTRUM CLASSIFIER

Landsat data is a collection of multispectral records of energy upwelling from terrain surfaces. Each terrain feature is represented by its own spectral distribution, or spectral signature. For most parametric classification algorithms, a large sample size is required to obtain an unbiased estimate of population parameters. A large sample size is not always available due to lack of reference data, high cost of sampling, or other factors. When there are sufficient laboratory measurements, or previous survey information for the area of concern, application of a correlation decision rule can help identify existing terrain features in the study area. The correlation decision rule may be described as follows:

If the n-dimensional vector for a single pixel of multispectral data is sufficiently correlated with the corresponding n-dimensional vector of an identified cover class, the pixel is assumed to belong to this class. The criterion for "sufficiently" may be varied to meet user needs, within the range of 0.000 to 1.000. If the n-dimensional vector meets the sufficiency criteria for more than one class, the pixel is assigned to the class with the highest correlation.

This correlation decision rule makes maximum use of the multispectral information for each target class. A classifier employing this decision rule, called Nearest Spectrum, was developed and tested under the following conditions:

1. with current training data used to train the classifier.
2. with previous information used to train the classifier.
3. with one pixel used to train the classifier.

Data Description

TM data for two different sites were selected for testing the reliability of the new classification algorithm. One was used for qualitative analysis of classification results since reference data were not available. The other was used for quantitative analysis of classification results using an enumerated reference data set.

Study area one is 3.6 km north/south and 3.24 km east/west, on Ross Island's Cape Crozier, Antarctica. The TM data used was from the 26 January 1985 acquisition for Path 055, Row 115. Figure 1-a is a feature plot for the study area, including two penguin rookeries (labelled R), based on a "manual" interpretation of a 3-band color composite prepared by M. R. Schwaller. Four land features - Rookery, Snow, Basalt, and Tuff - were identifiable. Snow, Basalt and Tuff have been grouped as "Land" on the plot for they were not targets of interest. Clouds (labelled C) and Sea Ice (labelled I) were also identifiable. While the boundaries of the rookeries may not be precise, their locations were correctly identified from the false-color composite and verified during ground investigation.

Study area two is a rectangular region 2.2 km north/south and 3.6 km east/west, in Leelanau County, MI. The TM data used were from the 11 July 1984 acquisition for Path 022, Row 029. Reference data for this site includes nine cover types enumerated from aerial photographs, a TM color composite, and field checks. These nine cover types are Water, Wetland, Conifers, Broadleaved Forest, Mixed Forest, Agriculture, Rangeland, Urban and Gravel Pit. A site specific accuracy obtained through pixel enumeration of these nine cover types was used to evaluate classification results.

Results and Discussion

Results of each of the three sets of the Nearest Spectrum classifier have been compared with results obtained with two other classifiers: a Nearest Mean classifier and a Gaussian Maximum Likelihood classifier.

Case One

Spectral separation of penguin rookeries from background materials results from the guano deposits in and around the penguin colonies. Deposits of guano and related debris made a unique organic soil which is distinguishable from surrounding areas in satellite images.

Figure 1-b is a result from the Nearest Spectrum classifier using conventional training sets, with only pixels classified as rookery shown on the plot. Some commission errors resulted from pixels of cloud classified as rookery. Two penguin rookeries were identified at a criterion of 0.000 (no minimum correlation requirement) for all targets.

Figure 1-c is a result from the Nearest Mean classifier using the same training sets, with all shaded areas pixels classified as rookery. The number of commission errors is high, with many pixels of cloud and non-rookery land features classified as rookery.

Figure 1-d is a result from the Gaussian Maximum Likelihood classifier using the same training sets. Although the two penguin rookeries were identified, the number of commission errors is very high. Not only have some pixels of cloud and other land features been classified as rookery, but also pixels on the edge of the sea ice were classified as rookery. The accuracy of penguin rookery classification may appear high due to the many commission errors, but these reduce confidence in the

classification results.

Figure 2-a is a result from the Nearest Spectrum classifier using the mean vector of the training sets for only the penguin rookery. Only pixels having a correlation coefficient between their vector and the group mean vector larger than 0.950 were classified as rookery. Two penguin rookeries were classified without any obvious commission errors, but some omission errors resulted from the high value placed on the correlation criterion.

The value of the correlation criterion can be varied to reflect the relative importance of omission and commission errors. Figures 2-b and 2-c are results from the Nearest Spectrum classifier using training sets for penguin rookery only, with criterion values of 0.900 and 0.850, respectively. The result obtained at 0.900 shows two rookeries identified with very few commission errors, a result similar to that in Figure 1-d from the Gaussian Maximum Likelihood classifier.

The great advantage of the Nearest Spectrum classifier is its simplicity of training set selection and quick target detection. Figure 2-d shows the result from the Nearest Spectrum classifier with only one pixel used as a training set, and the 0.950 criterion. This one training pixel was randomly selected from within the rookery and used for calculation of correlation coefficients with other pixels. Any other pixel whose correlation with the training pixel was less than 0.950 was classified as non-rookery. A comparison of Figure 2-a using mean values from larger training sets with Figure 2-d using single pixel training sets, both from the Nearest Spectrum classifier with a 0.950 criterion, show similar classification results. In those cases where large training sets cannot be identified, single pixels can be used as training sets with little loss in accuracy.

Varying the level of the correlation criterion permits the user to balance omission and commission errors. A large value of the criterion tends to reduce commission errors but increase omission errors, while a small value of the criterion tends to increase commission errors and reduce omission errors. Careful selection of the correlation criterion is essential in achieving desired classification results when using training information for only one target.

Since the Nearest Spectrum classifier makes maximum use of multispectral information, it has several advantages for target detection. Figures 1-b, 1-c, and 1-d show the Nearest Spectrum classifier to be superior to all others tested for detecting penguin rookeries.

Case Two

The second test was conducted for part of the Sleeping Bear Dunes Test Site in Leelanau County, MI, for which enumerated reference data area available. All nine cover types in the reference data were considered of equal importance. Site specific accuracy was obtained by comparing the classified map with the reference data pixel by pixel. Classification accuracies are shown in Table 1.

Column 1 shows results from the Nearest Spectrum classifier using training information from one random pixel for each target, except for Wetland and Mixed Forest whose spectral patterns were very close to those for Conifer and Broadleaved Forest. Overall accuracy was 80.1% with a Kappa coefficient of 66.0

Results in Column 2 are from the test using previous training information derived from an October 1982 TM data set for the same area. Overall accuracy was

76.6% with a Kappa coefficient of 61.6. Accuracy in classifying Water and Broadleaved Forest were particularly good. Without any current training data to modify previous training information, accuracy for Water was 98.5% and for Broadleaved Forest 92.4%.

Results in Column 3 were obtained with conventional sampling data (for 1984) as training information for each target and a 0.000 correlation criterion. An overall classification accuracy of 76.9% and a Kappa coefficient of 62.2 were obtained.

If different criteria were used for each target before correlation coefficients of all targets with an unknown pixel were compared, the Nearest Spectrum classifier would assign an unknown pixel to a class with unequal weights for each target. Results in Column 4 were from the Nearest Spectrum classifier with different criteria for each target (0.998 for Wetland and Mixed Forest, 0.000 for all others). Using the same conventional training data as before, an overall classification accuracy of 80.5% and a 66.9 Kappa coefficient were obtained. In order to increase classification accuracy, it is necessary to assign a higher criterion to classes with fewer pixels thereby preventing large commission errors for those classes.

Columns 5 and 6 show classification results from the Nearest Mean and Gaussian Maximum Likelihood classifiers respectively. Using the same conventional training data as with the Nearest Spectrum classifier, the Nearest Mean classifier yielded a 75.4% overall classification accuracy with a 61.6 Kappa coefficient, while the Gaussian Maximum Likelihood classifier yielded a 77.0% overall accuracy and a 65.2 Kappa coefficient.

Comparing classification results obtained for both overall classification accuracy and Kappa coefficient from the Nearest Spectrum, Nearest Mean and Gaussian Maximum Likelihood classifiers (bottom of Table 1: Columns 3 to 6) indicates that when the same training information is used, the Nearest Spectrum classifier gave better results than the Nearest Mean classifier and at least as good results as the Gaussian Maximum Likelihood classifier. Computational costs of the Nearest Spectrum classifier are close to those for the Nearest Mean classifier, and lower than those for the Gaussian Maximum Likelihood classifier.

Discussion

The Nearest Spectrum classifier offers an approach to supervised classification of multispectral data which differs from the Nearest Mean and Gaussian Maximum Likelihood classifiers. It assigns an unknown pixel to the classification category having the most similar spectral pattern -- the "Nearest Spectrum." Since the Nearest Spectrum classifier uses spectral patterns to classify unknown pixels, data with large spectral dimensions (many spectral channels) are desired. Data sets with few spectral dimensions do not provide as good classification accuracy unless the spectral patterns of the targets are very different in their smaller dimensional space. Another advantage of the Nearest Spectrum classifier is its ability to classify unknown pixels without being biased by the size of each cover type. If a target with very few pixels has a different spectral pattern from others, the target can be accurately classified.

The Nearest Spectrum classification algorithm should be an important addition to the current array of classifiers. It performs better with limited training information, has lower computational costs, and high classification accuracy. The calculation of correlation coefficients is much faster and easier than calculating density functions for a multivariate normal distribu-

tion. When the number of channels is larger than 7 (the number of TM spectral channels is currently 7, but various data transformations are often used to create "new" data channels), the Gaussian Maximum Likelihood classifier is ineffective due to a singularity of variance matrices, and the Nearest Mean classifier is disadvantaged by colinearity between multispectral channels. The Nearest Spectrum classifier can accommodate many spectral channels, and results tend to improve as the number of channels increases. Also, the Nearest Spectrum classifier is free of limitations based on size of training sets and is based on no statistical assumptions relating to estimation of population parameters.

Conclusion

The Nearest Spectrum classifier overcomes several difficulties in selection of training sets for a supervised classification of multispectral data. Either previous measurements or knowledge of current cover type can be used as input to train the classifier. It is free of statistical assumptions and has fewer limitations regarding training data. Thus, the Nearest Spectrum is a simple, efficient and relatively accurate supervised classification algorithm.

Literature Cited

- Chuvieco E. and R. G. Congalton, 1988. Using cluster analysis to improve the selection of training statistics in classifying remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, Vol. 54(9): 1275-1281.
- Hay, A. M., 1979. Sampling designs to test land-use map accuracy. *Photogrammetric Engineering and Remote Sensing*, Vol. 45(4): 529-533.
- Hixson, M., D. Scholz, N. Fuhs and T. Akiyama, 1980. Evaluation of several schemes for classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, Vol. 46(12): 1547-1553.
- James, M., 1985. Classification Algorithms. John Wiley & Sons, Inc., New York, NY, 209 pp.
- Justice, C. and J. Townshend, 1982. A comparison of supervised and unsupervised classification procedures on Landsat MSS data for an area of complex surface conditions in Basilicata, Southern Italy. *Remote Sensing of Environment*, Vol. 12: 407-420.
- Skidmore, A. K. and B. J. Turner, 1988. Forest mapping accuracies are improved using a supervised non-parametric classifier with SPOT data. *Photogrammetric Engineering and Remote Sensing*, Vol. 54(10): 1415-1421.
- Swain, P. H. and S. M. Davis, 1978. Remote Sensing: The Quantitative Approach. McGraw-Hill, Inc. New York, NY, 396 pp.
- Tom, C. H. and L. D. Miller, 1984. An automated land-use mapping comparison of the Bayesian Maximum Likelihood and Linear Discriminant Analysis algorithms. *Photogrammetric Engineering and Remote Sensing*, Vol. 50(2): 193-207.
- Wharton, S. W. and B. J. Turner, 1981. ICAP: An interactive cluster analysis procedure for analyzing remotely sensed data. *Remote Sensing of Environment*, Vol. 11: 279-293.

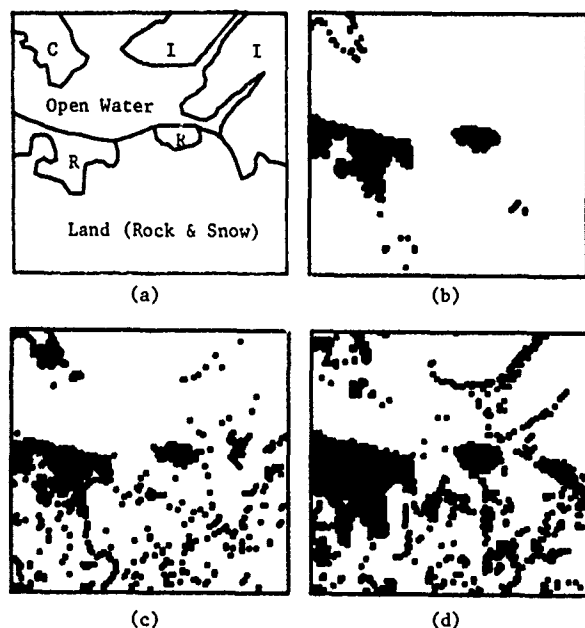


Figure 1. Comparison of classification results with different classifiers.

- (a) Cover types in the study area
 C = cloud I = sea ice
 R = rookery
 (b) Nearest Spectrum classifier
 (c) Nearest Mean classifier
 (d) Gaussian Maximum Likelihood classifier

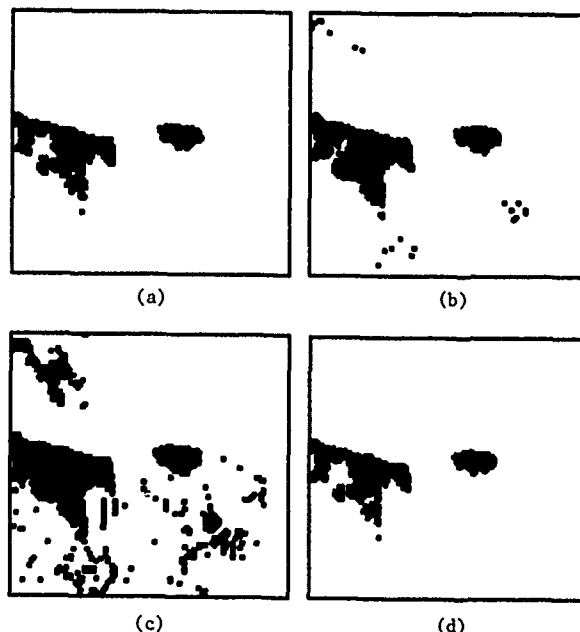


Figure 2. Results from the Nearest Spectrum classifier for one target class with different correlation criteria. (a), (b) and (c) based on conventional training sets; (d) with one pixel used as a training set.

- (a) Correlation criterion = 0.950
 (b) Correlation criterion = 0.900
 (c) Correlation criterion = 0.850
 (d) Correlation criterion = 0.950

Table 1. Summary of Classification Results Using Different Classifiers

| Category | Number of Pixels | Accuracy of Classification (%) | | | | | |
|---------------|------------------|--------------------------------|------|------|------|--------------|----------|
| | | Nearest Spectrum | | | | Nearest Mean | Gaussian |
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Water | 721 | 99.4 | 98.5 | 99.6 | 99.6 | 99.0 | 94.0 |
| Wetland | 50 | 0.0 | 22.0 | 26.0 | 6.0 | 40.0 | 64.0 |
| Conifer | 567 | 60.8 | 46.0 | 53.1 | 73.4 | 52.9 | 61.7 |
| Broadleaf | 4998 | 96.1 | 92.4 | 92.1 | 95.7 | 87.8 | 81.9 |
| Mixed Forests | 76 | 0.0 | 68.4 | 65.8 | 22.4 | 69.7 | 27.6 |
| Agriculture | 307 | 65.1 | 35.2 | 31.3 | 45.9 | 43.6 | 76.9 |
| Rangeland | 926 | 36.7 | 31.0 | 34.4 | 34.4 | 38.1 | 55.9 |
| Urban | 513 | 28.5 | 40.2 | 35.1 | 35.1 | 38.2 | 70.4 |
| Gravel Pit | 27 | 22.2 | 48.1 | 44.4 | 44.4 | 48.1 | 51.9 |
| Overall | 8185 | 80.1 | 76.6 | 76.9 | 80.5 | 75.4 | 77.0 |
| Kappa | | 66.0 | 61.6 | 62.2 | 66.9 | 61.6 | 65.2 |

1: one random pixel within each class. 2: previous training information
 3: current training information. 4: different criteria setting for each class.
 5: current training information 6: current training information

IMPROVED CLASSIFICATION USING IMAGERY FROM MULTIPLE SATELLITES

David S. Kauffman and Brian C. Robertson

MacDonald Dettwiler & Associates Ltd.
13800 Commerce Parkway, Richmond, B.C.
Canada V6V 2J3

ABSTRACT

The classification of surface cover types based on satellite-observed spectral properties is one of the primary applications of remotely sensed data, and has been a basic topic of interest since the launch of the first LANDSAT satellite. Identifying forest types, agricultural components, geological deposits, water sources and other environmental factors in an inexpensive and repeatable manner is crucial to the acceptance of remote sensing by the world community. Unfortunately, limitations of sensor resolution, spectral density and downlink bandwidth have restricted the amount of information received at the ground station and hampered automated image classification algorithms.

A major obstacle to the improvement of classification reliability from satellite images has been the limitation of analysing only a single multispectral scene in order to train and perform the classification. The reason for this limit is not cost of imagery, since the amount of time spent on acquiring ground truth, training the classifier and evaluating results is much higher. Rather, the inability to quantitatively compare data acquired at different times (multitemporal) or from different imaging sensors (multispectral) in either the geometric or radiometric domains has been the main impediment to using these additional information sources.

In order to perform analysis of multisensor and multitemporal imagery, research at MacDonald Dettwiler has focussed on the physical processes involved at each stage of the image acquisition and processing phase. Raw data from different satellites is geometrically registered and radiometrically calibrated to allow the overlay and subsequent quantitative comparison of the observed regions on the earth's surface. By carefully maintaining image integrity through the processing chain, the resulting output image pixels accurately represent the reflectance of the ground at map grid positions.

This paper explores the stages required for quantitative remote sensing using techniques of spacecraft modelling, sensor calibration, low-error resampling and atmospheric correction. Results of the process are illustrated on two independent images taken from the LANDSAT-5 TM sensor and SPOT MLA sensor several days apart. High radiometric and geometric registration between the corrected spectral signatures allowed training of one scene to be used to classify the other, in spite of different sensor characteristics and weather conditions under which the images were acquired.

Keywords: Classification, Geometric Registration, Resampling, Atmospheric Correction

1 REUSING CLASSIFICATION INFORMATION

The goal of classifying multispectral imagery is straightforward: *to map the digital pixel values to ground reflectance values*. Despite its apparent simplicity, this digital-number-to-ground-reflectance transformation is not a single function but is in fact the culmination of a complex set of links in the remote sensing processing chain. Establishing the weakest link, strengthening it, then reevaluating the chain is the ongoing process the remote sensing community is engaged in as we pursue the goal of quantitative remote sensing.

One measure of how well a system performs this $DN \rightarrow$ reflectance mapping is the reusability of classification training sets. Despite the fact that the acquisition and cataloging of ground cover is an expensive and time consuming activity, most training sets expressed in terms of spectral signatures cannot be reused on the same scene at any other time. Using "signature extension", MacDonald Dettwiler has utilized consistent approaches to geometric and radiometric correction to classify a SPOT MLA scene from the training set created for a Landsat TM scene.

2 QUANTITATIVE REMOTE SENSING

The notion of quantitative remote sensing implies the assessment of a physical quantity and error limits on each step in the processing chain. Both radiometric and geometric fidelity are required in order to establish both the ground reflectance at a point on the earth, as well as its location to some degree of accuracy.

The steps required to meet the stringent test of signature extension are the following:

1. Absolute Sensor Calibration
2. Precision Spacecraft Modelling
3. Atmospheric Modelling
4. Information-preserving Digital Resampling

3 SENSOR CALIBRATION

Sensor calibration establishes the relationship between radiant energy impinging upon each imaging element of the sensor and the measured and quantized value transmitted by the satellite.

For the Landsat TM sensor, pre-flight measurements were made to estimate the radiometric characteristics of the onboard lamps with an uncertainty of about 5% [Barker 85]. This information is supplemented by in-flight calibration data obtained by observing the calibration lamps after each swath. All this data has been incorporated into the absolute calibration step allowing accurate retrieval of measured radiance.

The SPOT satellite's CCD array have gains and dark current values that vary with time [CNES 85]. CNES updates these values several times a year providing up to date calibration information for ground processing stations.

Another form of calibration is from the White Sands, Nevada test area, where Dr. P. N. Slater has worked on simultaneous acquisition of radiance data while the satellite passes over, enabling the assessment of absolute calibration [Slater 87].

The end result of the absolute calibration phase is known as apparent radiance.

4 SPACECRAFT MODELLING

Radiometric accuracy is meaningless without the association of the spectral signature with a location on the earth's surface. MacDonald Dettwiler has used modelling of the physical imaging process since the early stages of the Landsat-IV program and now include Landsat-V, SPOT, and the Japanese MOS-1 [Erickson 89] satellite models. The specialization of modelling into

- sensor model,
- satellite model,
- earth model, and
- map model

have allowed the expansion of satellite types and enhancement of modelling to include long sections of the orbit track to a high degree using few control points [Sharpe 88].

5 DIGITAL RESAMPLING

Resampling is a complex, and sometimes misunderstood digital signal processing task in remote sensing. Done incorrectly, it can introduce artifacts and geometric distortion. We distinguish a *resampler* from an *interpolator* by the following criterion: A resampling operation preserves the mean and variance of the original sampled signal, while an interpolator is simply some method of establishing intermediate values between sampled points. By this definition, cubic convolution is an interpolator, but not a resampler. A resampler is built on an underlying model of how the signal was acquired and sampled, and reconstructs this underlying data prior to deriving a new set of sampled points. Wornell *et al.*, [Wornell 89] show that in order to preserve the radiometric integrity of the original imagery to within 1%, it is necessary to use at least an 8-point resampling kernel. Further, the resampling kernel should be derived such as to minimize the RMS (root-mean-square) difference between the original and resampled image.

| | Landsat TM | SPOT MLA |
|-------------|--------------|---------------|
| Date | Sept 5, 1987 | Sept 17, 1987 |
| Time | 18:27:58 GMT | 19:15:00 GMT |
| Sun Zenith | 47.0° | 47.7° |
| Sun Azimuth | 143.8° | 162.3° |
| View Zenith | 1.8° | 0.9° |
| Visibility | 20-30 miles | 12-15 miles |

Table 1: Image acquisition conditions for the Landsat TM and SPOT MLA scenes. Meteorological visibility was obtained from the Pacific Regional Office of the Atmospheric Environment Service.

As remotely-sensed imagery becomes widely used for quantitative analysis, machine analysis of resampled imagery has become commonplace. It is important that images that have passed through rigorous modelling and correction stages not be compromised by distortions introduced by the resampling operation.

6 ATMOSPHERIC CORRECTION

Atmospheric modelling and correction is the final phase before signature extension can be achieved. Although the previous stages establish and preserve the apparent radiance, estimating the actual reflectance of the surface still requires solar and atmospheric models.

MacDonald Dettwiler's work in atmospheric modelling has been described in [Robertson 89]. The analytic atmospheric model corrects for the effects due to sun position, gaseous absorption, radiation backscatter, direct and diffuse upward transmittance, and direct and diffuse downward transmittance.

Upon completion of the atmospheric correction stage, the data has been converted from apparent radiance to corrected reflectance, that is an accurate estimate of the spectral reflectance properties of the surface integrated over a region on the ground.

7 SIGNATURE EXTENSION

The high degree of spectral overlap between the SPOT HRV multi-spectral bands and Landsat TM bands 2, 3 and 4 makes signature extension possible. Even so, care must be taken to account for the differences in the spectral response of the two sensors. Since this information is integrated in the calibration phase described above, the reflectance data generated by these procedures are directly comparable.

Two scenes were acquired for the signature extension test; a Landsat TM image acquired September 5th, 1987, and a SPOT MLA image from September 17th. The acquisition dates were near enough in time that the ground cover was effectively unchanged between the two images, yet significant differences in the atmospheric conditions enabled the assessment of the atmospheric modelling and correction procedures. Table 1 shows the conditions under which the images were acquired.

To assess the ability to extend spectral signatures between the two datasets, rectangular regions of uniform appearance were selected from the geocoded imagery. All sites cover the same ground area, hence there were four times as many SPOT pixels at a geocoded 12.5 metre resolution than TM pixels at a geocoded 25 metre resolution. Figure 1 shows a graph of the raw digital numbers at the test sites. Regression analysis shows that there is a 0.98 correlation between each band pair, which indicates the sites chosen truly were stable

TM vs MLA Raw Digital Numbers
For Homogenous Sites in Vancouver Datasets

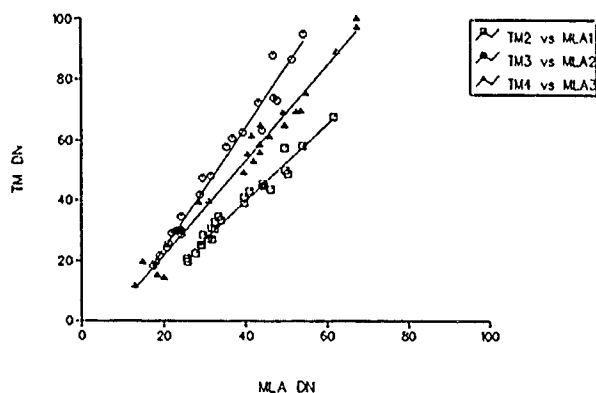


Figure 1: Comparison of the raw digital numbers at 24 test sites in the TM and MLA imagery of Vancouver. While there is a strong linear relationship, no absolute comparisons may be made between the TM and MLA data as there is no way of relating the raw DNs to surface properties. Time varying sensor properties and environmental conditions mean the slope and intercept of the line relating the TM-MLA band pairs changes from scene to scene.

over the 12 day period. The fact that the slopes and intercepts of lines fit through the data are different in each band, and far different than an ideal linear relation with a slope of one and a zero intercept highlights the problems associated with the quantitative comparison of multitemporal multisensor datasets.

Figure 2 shows the apparent and corrected reflectance of the test sites for the TM band 4 and MLA band 3 data. The relationship between the corrected reflectance values is close to the ideal relationship with a near one slope, and near zero intercept. Once sensor-specific effects are removed, such as calibration and normalization of spectral response, and by applying the appropriate atmospheric model to remove atmospheric effects, the correspondence between the two images is within 0.02 reflectance units in all bands. The remaining differences, about 6 TM digital values, are well within the absolute calibration accuracies quoted for the TM and MLA sensors.

It is important to consider that the results obtained are without resort to relative operators such as histogram equalization. Instead, each image has been independently modelled and corrected to an absolute scale of reference, the essence of quantitative remote sensing.

8 CONCLUSIONS

Using analytic techniques, we have modelled the radiometric and geometric acquisition process for satellite imagery, including the sensor characteristics, satellite motion, earth and map models, and atmospheric conditions. A systematic test of applying these methods to two images from different satellites shows the restoration of sampled energy at the satellite to ground reflectance, and in doing so removes most satellite dependencies, producing imagery with less dependence on satellite type, time of day, and atmospheric conditions.

In particular, the ability to normalize imagery from different satellites such that their estimates of ground reflectance is within 10% enables a major lowering of the cost of acquiring ground reflectance data directly by allowing the reuse of information through signature extension.

Apparent and Corrected Reflectance
For Homogeneous Areas in Vancouver Datasets
Landsat TM4 vs SPOT MLA3

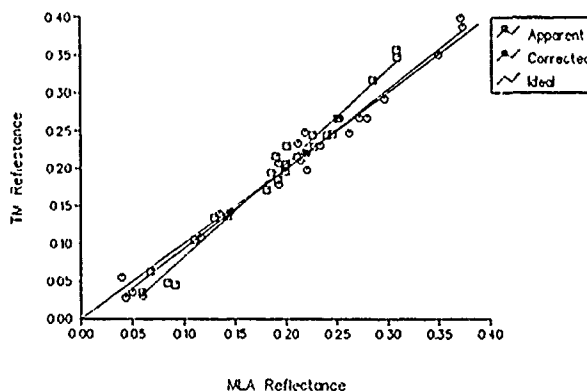


Figure 2: Comparison of apparent and corrected reflectance at the 24 test sites in TM band 4 and MLA Band 3 imagery. Notice that the corrected TM and MLA reflectance values may be directly compared as they are represented on the same scale, and the line fit through the data has a slope close to the one and a near-zero intercept.

References

- [Barker 85] J. L. Barker, D. L. Ball, K. C. Leung, and J. A. Walker. Prelaunch absolute radiometric calibration of the reflective bands on the Landsat-4 photoflight Thematic Mapper. In *Landsat-4 Science Characterization Early Results*, pages II-277-372, Goddard Space Flight Center, 1985.
- [CNES 85] CNES. *Spot to direct receiving station interface document*. Technical Report S-IF-O/E-10-CN, Centr National d'Etudes Spatiales, 1985.
- [Erickson 89] Arvon Erickson, Brian Robertson and Kelly Wiebe. Geometric and radiometric correction of MOS-1 imagery in a Canadian processing system. In *IGARSS Proceedings, IEEE International Geoscience and Remote Sensing*, July 1989.
- [Robertson 89] Brian Robertson and P. M. Teillet. Integrated atmospheric correction of multispectral satellite imagery. In *IGARSS Proceedings, IEEE International Geoscience and Remote Sensing*, July 1989.
- [Slater 87] P. N. Slater, S. F. Biggar, R. G. Holm, R. D. Jackson, Y. Mao, M. S. Moran, J. M. Palmer, and B. Yaun. Reflectance and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sensing of Environment*, 22:1 37, 1987.
- [Sharpe 88] Bruce Sharpe and Kelly Wiebe. Planimetric accuracy in satellite mapping. In *Proceedings of the XVI International Congress of the ISPRS*, 1988.
- [Wornell 89] Gregory W. Wornell, David S. Kauffman, and Bruce Sharpe. Minimum mean-square error resampling for remotely-sensed imagery. In *IGARSS Proceedings, IEEE International Geoscience and Remote Sensing*, July 1989.

CLASSIFICATION OF SATELLITE IMAGES USING CONTEXTUAL CLASSIFIERS

F. GONZALEZ ALONSO and S. LOPEZ SORIA

Sección de Teledetección Agraria. Instituto Nacional de
Investigaciones Agrarias (INIA). Apdo. 8111. Madrid 28080. España.

Abstract

The classification accuracy obtained from the classification of satellite images using pixel by pixel conventional methods could be improved if the contextual information is considered jointly with the spectral information in the same strategy of classification. A computer program has been developed to implement a contextual classifier algorithm. The accuracy improvement was evaluated and tested in two pilot zones of Central Spain.

1.- INTRODUCTION

The digital images coming from the earth observation satellites (Landsat TM and MSS, SPOT, NOAA, MOSS), always contain three types of information:

Firstly, a spectral information resulting from the interaction between the electromagnetic energy and the materials that can be found on the earth surface.

Secondly, a temporal information derived of the satellite orbits repetitivity that allows to study the evolution of the spectral signatures of the surfaces and makes it possible to monitor their temporal changes.

Finally, there is a spatial information, that represents the spatial organization of pixels that make up the image obtained by the sensor.

Traditionally the first two types of information have been the most frequently used in the application of different techniques and image processing methodologies, specially when the objective has been to classify the remote sensing images by supervised or not supervised procedures.

However, the spatial information has been less employed and the majority of classifiers developed have been exclusively based on the application of one pixel by pixel strategy that does not take into account the possible relations or similarities that may exist among one pixel and their neighbours.

The main reason for not using the spatial information is that the spectral information can be easily analysed pixel by pixel whereas

the use of the spatial information normally involves the simultaneous consideration of several pixels during all the computational process.

The results of correct classification obtained from the classification of satellite images by conventional methods can be improved if the spatial information is considered jointly with the spectral information in the same strategy of classification (Swain *et al.*, 1979).

2.- SPATIAL INFORMATION AND CONTEXTUAL CLASSIFIERS

The spatial information contained in the satellite images can be subdivided into two types: texture and context.

According to Gurney and Townshend (1983) texture refers to a description of the spatial variability of tones within part of a scene.

Whereas texture refers to the spatial variation within a contiguous group of pixels the context of a pixel (or a group of pixels) refers to its spatial relationships with pixels in the remainder of the scene. Thus, contextual classification of any pixel can, potentially at least, involve the use of any other pixel or group of pixels from throughout the whole scene (Gurney and Townshend, 1983).

We can briefly define the contextual classifier as those image classification techniques that consider simultaneously the spectral and spatial characteristics of the images to achieve more accurate classification results.

3.- METHODOLOGY

One intuitive way of incorporating spatial information into the classification strategy is to hypothesize that the ground cover type of given pixel is not independent of the ground cover types of its neighbouring pixels.

In this way, it is more likely that a wheat parcel grows in the neighbourhood of a barley parcel than in the middle of a high density urban area.

From a statistical classification standpoint we may have a better chance of correctly classifying a given pixel if we consider not only the spectral data associates with the pixel itself but as well the spectral data and ground cover classes of its neighbours.

The methodology adopted to accomplish this research project was proposed by Swain *et al.*, (1979) and Siegel *et al.*, (1980) and can be summarized in the following way:

The image data to be classified are assumed to be a two dimensional I by J array of multivariate pixels.

Wi will consider as the p-array the pixel that must be classified and the p-1 neighbour pixels.

In our case p=3 and when we try to classify the pixel (i,j) we also consider the pixels (i,j-1) and (i,j+1). Associated with the pixel at "row i" and "column j" is the multivariate measurement n-vector $x_{ij} \in R^n$ and the true class of the pixel $\theta_{ij} \in \Omega = \{\omega_1, \dots, \omega_C\}$

The measurements have class-conditional densities $f(x|\omega_k)$, $k=1,2,\dots,c$ and are assumed to be class ω_k conditionally independent.

When context is ignored the classification of the pixel in the (i,j) position depends only on the measurement vector of the pixel to be classified.

As we want to incorporate some neighbourhood information in the decision process, let x_{ij} be a p-vector of measurement vectors associated with pixel (i,j) to be classified and let θ_{ij} be the corresponding p-vector of actual θ_{ij} classes.

As indicated by Siegel *et al.*, (1980), for each pixel, for each class in Ω a discriminant function g is calculated. The pixel is assigned to the class for which g is greater.

The expresion of g for pixel (i,j) being in class ω_k is:

$$g_k(x_{ij}) = \sum_{\theta_{ij} \in \Omega} \left[\prod_{l=1}^p f(x_l | \theta_l) \right] G^p(\theta_{ij})$$

where:

$x_l \in x_{ij}$ is the measurement vector from the lth pixel in the p-array (for pixel (i,j))

$\theta_l \in \theta_{ij}$ is the class of the lth pixel in the p-array (for pixel (i,j))

$f(x_l | \theta_l)$ is the class-conditional density of x_l given that the lth pixel is from class θ_l

$G^p(\theta_{ij}) = G^p(\theta_1, \theta_2, \dots, \theta_p)$ is the a priori probability of observing the p-array $\theta_1, \theta_2, \dots, \theta_p$

G^p is referred to as the context distribution and can be evaluated from a previous classification of the scene

4.- APPLICATION AND RESULTS

To apply practically the methodology previously exposed we have developed the

program CONTEXT written in FORTRAN language and implemented in a CDC CYBER 180 computer.

This program was applied in several pilot zones with different agricultural landscapes in the Toledo province (Central Spain). The pilot zones were previously classified using a Landsat-4 MSS image and a pixel by pixel supervised procedure (González *et al.*, 1988).

Figure 1 represents the classification results obtained in the pilot zone of Cazalegas reservoir following a pixel by pixel supervised classification procedure Fig.1 A and a contextual classifier Fig.1 B.

In order to analyze the homogeneity effect produced by the contextual classifier we have defined two indices to measure the homogeneity of a zone.

Diversity index ID:
ID=CD/CP

where

CD=number of different combinations present in image composed by three pixels.
CP=theoretical number of possible different combinations.

In our case as we have 13 different classes of ground cover CP=13**3=2197.

Homogeneity index IH.
$$IH = \frac{\sum_{i=1}^{13} G(i,i,i)}{NP}$$

where

$G(i,i,i)$ =number of pixels with pure context in each class.
NP= number of pixels in the scene.

| | ID | IH |
|-------------------------------|--------|--------|
| Classification pixel by pixel | 0.1520 | 32.8 % |
| Contextual classifier | 0.0864 | 47.5 % |

ID=Diversity Index
IH=Homogeneity Index

Table 1. Results for ID and IH in the Cazalegas reservoir pilot zone

From Table 1 we can deduce that the application of the contextual classifier decreases the diversity (43.15%) and increases the homogeneity (44.81%) in the classification obtained.

Another way of analyzing the utility of the contextual classifier is to generate a simulated image from the original classification and then classify this simulated image following the pixel by pixel and contextual classifier procedures. After calculating the percentage of coincidence of these two classifications with the original classification we can evaluate the accuracy improvement resulting from the use of the contextual classifier.

In the case of Cazalegas reservoir pilot zone these percentages were 48.26 and 50.15 for the pixel by pixel, and contextual classifier procedures respectively.

In this particular case it seems that the context strategy is not very significant. This is probably due to the existing interaction between the environment and the

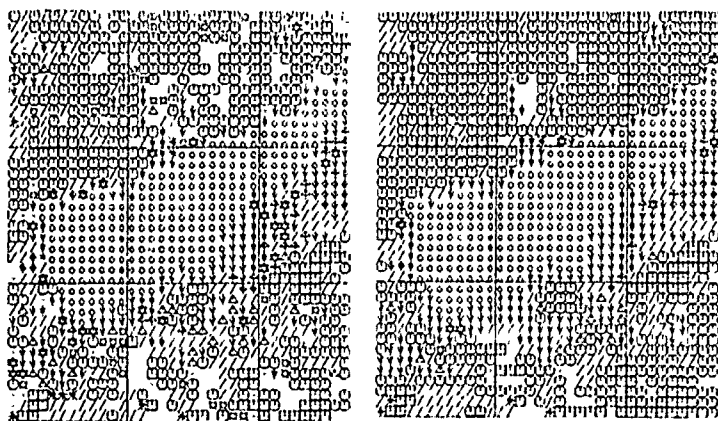


Fig. 1: Classification results obtained in the pilot zone of Cazalegas reservoir following a pixel by pixel supervised classification procedure (fig.1 A) and contextual classifier (fig.1 B).

spatial resolution of the sensor (80m. in this case) is not sufficient to produce a high spatial autocorrelation in the image that allows to the contextual classifier to exploit this circumstance in a suitable way.

Another example (olive trees plantation) is showed in Fig.2

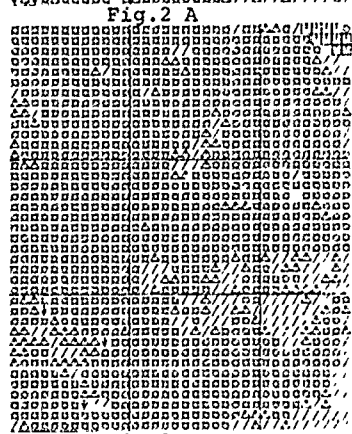
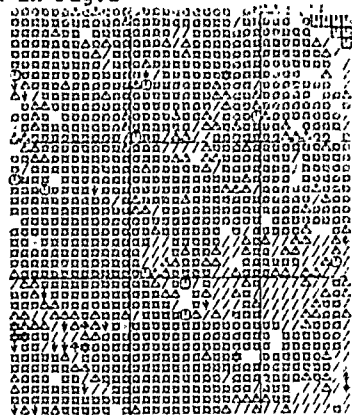


Fig.2 B

Fig.2: Classification results obtained in a olive trees pilot zone following a pixel by pixel supervised classification procedure (fig.2 A) and a contextual classifier (fig. 2 B).

- / CEREALS
- * LEGUMINOUS
- OLIVE TREE PLANTATIONS
- ☆ GREEN OAK FORESTS
- PLOUGHED SOILS
- OPEN GREEN OAK FOREST 1
- OPEN GREEN OAK FOREST 2
- ↑ PINUS PINEA
- ◇ CHESTNUTS
- △ PINUS PINASTER PLANTATIONS
- ▼ CISTUS SSP
- ▼ SHRUBS
- RESERVOIRS

| | ID | IH |
|-------------------------------|--------|--------|
| Classification pixel by pixel | 0.0605 | 54.0 % |
| Contextual classifier | 0.0245 | 70.6 % |

ID=Diversity Index
IH=Homogeneity Index

Table 2. Result for ID and IH in the olive trees pilot zone.

As we can see in Table 2 the olive trees zone is less diverse than the Cazalegas reservoir pilot zone. The application of the contextual classifier in the olive trees pilot zone procedure a decrease in the ID index of 59.50% and an increase in the IH index of 30.74%.

After the application of the contextual classifier we must note that the olive trees zone is 3.52 times less diverse and 1.48 times more homogeneous than the Cazalegas reservoir pilot zone, partly due to the higher intrinsic homogeneity in the case of the olive trees zone.

The generation and classification of simulated images also was done in the case of the olive trees pilot zone and the percentage of coincidence of the pixel by pixel classification and contextual classification of the simulated image with the original classification were 12.14 and 70.36 respectively.

The previous results show, in the first place, the high number of errors obtained in the simulated image of this zone, by the high spectral similarity existent in the classification of reference between the classes "ploughed soils" and "olive tree plantations". The consequence of this circumstance is that when we make the classification of the simulated image by the pixel by pixel procedure, their percentage of coincidence with the original classification is very small.

Secondly, we can also observe that in spite of the deficient radiometric simulation, the classification results can be substantially improved if we use the contextual classifier.

In this case, the percentage of coincidence was improved from 12.14 to 70.36. In other words, the percentage of correct classification was increased by 58.22 if we use the contextual classifier versus the pixel by pixel classifier.

The previous considerations are showed in Figure 3. A possible explanation for these results could be that as the surface of the olive tree plantations is normally bigger than the spatial resolution size of the MSS sensor, we can expect the existence of a high spatial autocorrelation among the different neighbouring pixels and, in these circumstances it can be very interesting to incorporate the context information at the classification strategy.

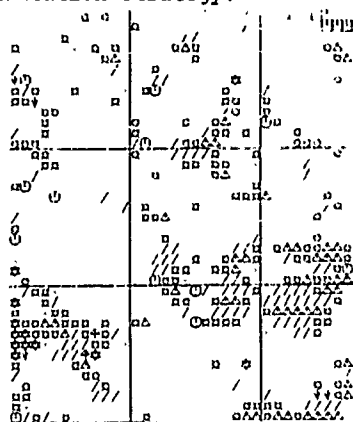


Fig.3 A

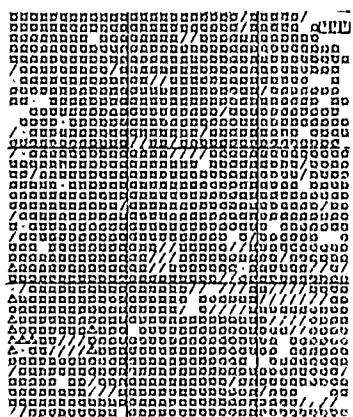


Fig.3 B

5.- CONCLUSIONS

The use of contextual classifiers produce classification results more homogeneous than those obtained when a pixel by pixel classification procedure are employed.

This fact was observed in different landscapes of Central Spain. The adoption of a strategy that considers in the classification process the spatial information contained in the digital images, improves the percentages of correct classification, specially if the images have a high spatial autocorrelation derived from the interaction between the type of landscape in the image and the spatial resolution of the sensor.

The use of the contextual classifier was specially interesting for the classification of olive trees homogeneous zones.

The contextual classifiers normally require large amounts of computation time and this circumstance can be a serious inconvenient if we try to use them over large areas.

Finally, this kind of classifiers allows a better description of the context in a scene and can be very useful for the classification of high resolution data as it is the case with satellite images coming from the TM and HRV sensors.

REFERENCES

- Gonzalez, F., Cuevas, J., and Moro, J., 1988. Aplicación del Sistema Informático ERAFIS al reconocimiento y discriminación de los usos del suelo en la Provincia de Toledo a partir de imágenes digitales MSS-Landsat 4. Instituto Nacional de Investigaciones Agrarias, Madrid. España.
- Gurney, C.M., and Townshend, J.R.G., 1983. The use of contextual information in the classification of remotely sensed data. *Photogram. Eng. Remote Sensing*, 49, 55.
- Siegel, H.J., Swain, P.H., and Smith, B.W. 1980. Parallel processing implementations of a contextual classifier for multispectral remote sensing data. *Proc. of the 1980 Machine Processing of Remotely Sensed Data Symposium*, Purdue, June, 19.
- Swain, P.H., Siegel, H. and Smith, B.W. 1979. A method for classifying multispectral remote sensing data using context. *Proc. of 1979 Machine Processing of Remotely Sensed Data Symposium*, Purdue, June, 343.

Fig. 3: Classification results obtained for a simulated image of the olive trees pilot zone following a pixel by pixel supervised classification procedure (fig.3A) and contextual classifier (fig.3B).

CLASSIFICATION OF SPOT IMAGE USING SPECTRAL AND SPATIAL FEATURES OF PRIMITIVE REGIONS WITH NEARLY UNIFORM COLOR

T. Kusaka, H. Egawa and Y. Kwata

Kanazawa Institute of Technology, Ishikawa 921, Japan

ABSTRACT

This paper describes how the SPOT image is classified by spectral and spatial features of small regions (primitive regions) with nearly constant color. At first, the SPOT image is segmented into the primitive regions using an edge-based segmentation technique. Next, spectral and spatial properties such as average grey level in each spectral band (AL), normalized vegetation index (NVI), variance of brightness (VB) and form factor (F) are extracted for each primitive region. Then, the statistical quantities for AL, NVI, VB and F are computed for the classes corresponding to typical land cover types in the study site. Finally, all the regions are classified by using the estimated class statistics. Consequently, it is found that the present method yields better classification results than those obtained from the classification using only multispectral features.

Keywords: Classification of SPOT Image, Spectral and Spatial Feature, Region Segmentation, Color Edges, Primitive Region.

1. INTRODUCTION

The statistical methods based on multispectral classifiers are still widely used in the field of remote sensing. However, in the statistical classification analysis of high resolution satellite images such as Landsat TM and SPOT images, it is difficult to recognize complex surface patterns reliably. This is because spatial features that characterize the desired objects and structures are not taken into account. In fact, the color composite images of such high resolution satellite data visually give us many details about scene structures. The substantial improvements in classification performance of high resolution satellite images will require new techniques.

In this paper we describe how the SPOT image is classified by spectral and spatial features of small regions with nearly constant color. The present method is based on the assumption that color edge points are very important features in the analysis of the complex natural scene. The color edge points generated from three spectral images separate the regions of different color, but there are small gaps in edge points. In the present study the completely enclosed regions are extracted by using isograms of equal distances from the

boundary points of partially enclosed regions. The small enclosed regions with nearly uniform color will be referred to as the primitive regions. Then, spectral and spatial features such as the average grey level in each spectral band (AL), normalized vegetation index (NVI), variance of brightness (VB), area (A) and form factor (F) are computed for the primitive regions corresponding to various land cover types (rice field, urban area, residential area, bare soil such as playground, forest, highway and river). Finally, all the primitive regions are classified by the spectral and spatial features of primitive regions described above.

2. REGION EXTRACTION METHOD

The region extraction methods using edge-based segmentation techniques have not been successful because there are small gaps in the extracted edge points. The method used here is a slight extension of that described in Ref.1 (Egawa, 1988), and is such a segmentation method that the regions are correctly separated, despite of small gaps in edge points. We use the image of Fig.1 to illustrate the method described in the present study. (Fig.1 is actually the false color image of SPOT data taken on Aug. 20, 1986 in which the XS1 channel is assigned to Blue, XS2 to Green, and XS3 to Red.) In Fig.1, the bright regions in the central portion of the image represent the urban and suburban area. The dark regions in the left-hand side of the image represent the agricultural land. The woodland is located at the lower right portion.

2.1 Detection of color edges

Although it is difficult to detect color edges from multispectral images, we consider that the color edges appears at the portions where the brightness steeply changes in the image (Nevatia, 1982). Therefore, the Sobel operator approximated by the absolute-value operator was applied to produce edge magnitudes at every pixel in each spectral image. Then, the color edge magnitude M at a pixel was defined as follows:

$$M = \sum_{i=1}^3 W(i) \times G(i) \quad (1)$$

where $w(i)$ and $G(i)$ are the weighting function and the edge magnitude in spectral band i , respectively. For simplicity, we chose as $w(i)=1/6$.

After the values of M have been computed at all the pixels of the image, the resulting edge magnitude data are thresholded at an appropriate value to produce

sharp ridges. The threshold T was selected at the median value between the average color edge magnitude and the color edge magnitude corresponding to the highest peak in the histogram generated from all values of M . Then, we performed the following operation to leave only the real edge points.

If the color edge magnitude M at a given point (i,j) is greater than T , then the magnitudes of its two neighbors are examined in vertical, horizontal and two diagonal directions. If M is greater than each value of two neighbors in one or more direction of the four, that point is determined to be the real edge point, and the value 1 is assigned to the corresponding point of another array, $EG(i,j)$. Otherwise, the value of 0 is set to $EG(i,j)$.

Fig.2 shows the thinned color edges. In Fig.2, the isolated edges that consist of one or two edge points are eliminated.

2.2 Extraction of primitive regions

The thinned edge points do not form the completely enclosed regions since there may be small gaps in edge points.

In the present study we extract small enclosed regions in the similar way that Harms, et al. (1986) have used the distance-transform algorithm (Rosenfeld, 1976) for the blood cell analysis. The work described here differs from that described in Ref.3 (Harms, 1986) by the method of separating regions.

The following operations were performed in extracting primitive regions (small enclosed regions) :

(1) In the array EG , all edge points are set to 0 and others to 1, and then each pixel is expanded to a 2×2 pixels in order that small regions are not absorbed to the regions that surround them.

(2) Isograms of equal distances from edge points are constructed, and local isogram maxima whose values are greater than 1 are found. They are regarded as the nuclei of primitive regions.

(3) Each nucleus is labelled and expanded to separate the enclosed regions.

After steps (1) through (3) have been applied to every point of the array EG , the edge pixel is merged to its adjacent regions. However, it is likely that only the edge points form a new region for the thin elongated structures such as road and small river. Therefore, we used the color contrast between the edge point and its adjacent regions to determine the edge pixels to be merged. The color contrast C is defined as follow:

$$C = \sum_{j=1}^3 |I(j) - \bar{I}(j)| \quad (2)$$

where $I(j)$ and $\bar{I}(j)$ are the grey level of the edge pixel and the average grey level of its adjacent primitive region in band j , respectively. The values of C are computed for all primitive regions that are adjacent to an edge pixel, and the primitive region with minimum contrast C_{min} is found. If C_{min} is less than T , where T is the threshold value described in the previous section, then the edge pixel is merged to the region with minimum contrast.

The size and number of primitive regions after merging edge points are summarized in Table 1. We can see from Table 1 that there are 14421 regions in the segmented image which consists of 262144 pixels. Fig.3 shows the final results obtained from our region extraction method. (Actually, Fig.3 is the color composite image.)

3. CLASSIFICATION OF SPOT IMAGE

3.1 Properties of primitive regions

We use the spectral and spatial features of primitive regions to classify all the regions in SPOT image. The following properties for each primitive region were extracted:

(A) Spectral properties

a) average grey level in each spectral band (AL)

b) normalized vegetation index (NVI)

(B) Spatial properties

c) area (number of pixels: A)

d) size (maximum distance: D)

e) form factor (F)

(C) Texture measure

f) variance of brightness (VB)

The NVI is defined as $(XS3 - XS2) / (XS3 + XS2)$, where $XS2$ and $XS3$ are average grey levels of a primitive region in band 2 and band 3, respectively. D is the maximum value of isograms of equal distances in each region. The form factor F is defined as A/D^2 , and becomes nearly equal to unity for the square shape regions. For elongated structures we have large values of F .

Property f) is considered to be the texture measure of a primitive region because VB provides the low value for homogeneous regions and the high value for highly textured regions. Here, the brightness is defined as the value of second principal component, because the components of its eigenvector yield all the positive values in the principal component analysis of the present study site.

3.2 Classification method

The classification method used here is as follows:

(1) The regions with 100(1-)% confidence are constructed in a three dimensional feature space by using the statistical quantities for values of AL alone in classes corresponding to typical land cover types. Since the decision is made by comparing the squared distances (Mahalanobis distances) to each class center, we can expect that a sample point belongs to the class k if $MD(k) < \chi^2_k(3)$, where $MD(k)$ is the Mahalanobis distance between that point and the center of class k , and $\chi^2_k(3)$ is a χ^2 variate with 3 degrees of freedom. In this study $\alpha=0.1$ is used.

(2) If a sample point is assigned to the overlap domains among classes, then the values of NVI, VB, D and F are used to decide to which class that point belongs. In this case, we also use the class statistics for NVI, VB, D and F to evaluate the Mahalanobis distances from the sample point to each center of the overlapping classes. The sample point is finally assigned to the class that minimizes the Mahalanobis distances.

It should be noticed that the properties of primitive regions: NVI, VB, D and F are used to discriminate the regions with almost the same spectral pattern.

3.3 Classification results

A wide variety of land cover types is present in the study site. The legend of the major land cover types for the image under consideration is listed in Table 2.

In order to estimate classification accuracies, the 1180 pixels were randomly selected from the whole image (512x512 pixels area). At each of these points, the land cover type as shown in Table 2 was determined by the interpretation of 1:25,000 land use maps, aerial photographs and the ground truth data provided by the joint work of Geographical Survey Institute (GSI) and Ishikawa Research Laboratory for Public Health and Environment (IRLPHE) (Ueno, et al., 1985). The 560 points were also selected from this dataset to obtain the statistical quantities for each of the classes corresponding to cover types specified above.

To compare with the results of the multispectral pattern classification (the classification using multispectral data of each pixel) and those of the region-based classification, we classified the whole image in the study site in terms of only spectral values of primitive regions, i.e., the values of AL. We used a simple decision rule. In other words, a sample point of interest is classified into the class which minimizes the Mahalanobis distance between the sample point and the class mean. The classification results are summarized in Table 3 and 4. In Table 3 the accuracy of the classification using multispectral features of each pixel is shown, whereas the accuracy of the classification using the AL-values of primitive regions is listed in Table 4. In Table 3 and 4, the classification accuracy of a given land cover type is defined as the ratio of the number of pixels corresponding to the cover type as classified on the SPOT image to that belonging to the cover type as predetermined on the ground truth data. The multispectral pattern classification yields the accuracy of 82 % for all the cover types listed in Table 3. On the other hand, the classification using spectral data of primitive regions yields the accuracy of 86 %, an improvement over the accuracy of the multispectral pattern classification.

Now, we present the classification results using both spectral and spatial features of primitive regions. The classification method described in the previous section was applied to the whole image in the study area. The results are shown in Table 5. The classification yields the accuracy of 89% for all land cover types. This indicates that the classification method using both spectral and spatial features of primitive regions significantly improves the classification accuracy.

4. CONCLUSIONS

The SPOT image was segmented into small regions with nearly constant color using an edge-based segmentation technique. The segmented small regions were classified by spectral and spatial features of them. Consequently, it is found that the classification method significantly improves the accuracy in classification performance, in comparison with the results obtained from the classification method using only multispectral features. Thus, the region segmentation using color edge points can be a powerful and effective aid to improve the classification accuracy.

REFERENCES

1. Egawa, H and T. Kusaka, Region extraction in SPOT data, Proc. of ISPRS, Vol. 27, part b7, Commission VII, pp. 142-151, 1988.
Egawa, H and T. Kusaka, Region extraction in SPOT data, Geocarto International, Vol. 3, pp. 25-30, 1988.
2. Nevatia, R, Machine perception, Prentice-Hall, Inc., 1982.
3. Harms, H, U. Gunzer and H. M. Aus, Combined local color and texture analysis of stained cells, Computer vision, graphics, and image processing, Vol. 33, pp. 364-376, 1986.
4. Rosenfeld, A and A. C. Kak, Digital picture processing, Academic Press, 1976.
5. Ueno, S, Y. Kawata and T. Kusaka, Optimum classification of Landsat Thematic Mapper data for ecological study, Proc. of 19th Int. Symp. on Remote Sens. of Environ., ERIM, pp. 21-25, 1985.



Fig.1 Original SPOT image in the study area



Fig.2 Thinned edge image

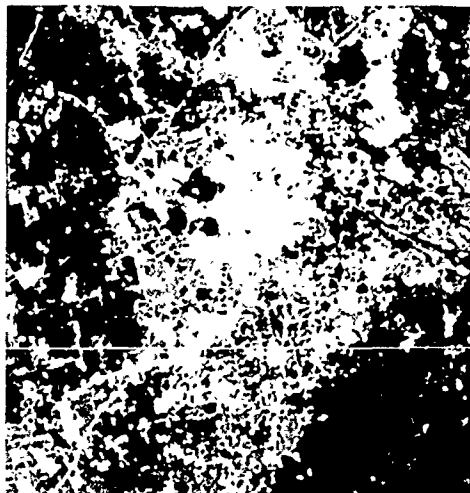


Fig.3 Final results obtained from the region extraction processing

Table 1 The size and the number of primitive regions after merging edge points

| Size (Number of pixels) | Number of primitive region |
|----------------------------|----------------------------|
| 1 ~ 5 | 3 2 1 1 (22.2%) |
| 6 ~ 10 | 4 3 4 6 (30.0%) |
| 11 ~ 15 | 2 5 1 1 (17.3%) |
| 16 ~ 20 | 1 3 8 7 (9.6%) |
| 21 ~ 25 | 8 0 6 (5.6%) |
| 26 ~ 30 | 5 2 4 (3.6%) |
| 31 ~ 35 | 3 2 7 (2.3%) |
| 36 ~ 40 | 2 4 3 (1.7%) |
| 41 ~ 45 | 1 8 2 (1.3%) |
| 46 ~ | 9 4 4 (6.5%) |
| Total number | |
| 1 4 4 8 1 | |

Table 2 Legend of land cover types in the study site

| class number | cover types |
|--------------|--|
| 1 | urban area |
| 2 | agricultural land (rice field) |
| 3 | residential area (thickly housed area) |
| 4 | playground |
| 5 | forest |
| 6 | highway |
| 7 | river |

Table 3 Accuracy(%) of classification using multispectral data of each pixel

| | I | II | III | IV | V | VI | VII |
|---|------|-------|------|-------|-------|------|------|
| 1 | 64.8 | 0.5 | 6.0 | 0.5 | 0.0 | 23.1 | 5.0 |
| 2 | 0.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 9.0 | 4.1 | 72.4 | 0.0 | 1.1 | 10.4 | 3.0 |
| 4 | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 | 0.0 |
| 6 | 42.7 | 0.0 | 16.9 | 0.0 | 0.0 | 40.3 | 0.0 |
| 7 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 99.0 |

Table 4 Accuracy(%) of classification using only AL-values of primitive regions

| | I | II | III | IV | V | VI | VII |
|---|------|-------|------|------|------|------|------|
| 1 | 62.4 | 0.0 | 2.9 | 0.8 | 0.0 | 28.9 | 5.0 |
| 2 | 0.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 9.3 | 6.0 | 67.3 | 0.0 | 0.0 | 17.1 | 0.4 |
| 4 | 1.3 | 0.0 | 0.0 | 98.7 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 0.6 | 0.6 | 0.0 | 98.8 | 0.0 | 0.0 |
| 6 | 48.0 | 0.0 | 7.1 | 0.0 | 0.0 | 44.9 | 0.0 |
| 7 | 0.0 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 | 73.0 |

Table 5 Accuracy(%) of classification using both spectral and spatial features of primitive regions

| | I | II | III | IV | V | VI | VII |
|---|------|-------|------|------|------|------|------|
| 1 | 66.9 | 0.0 | 2.9 | 0.8 | 0.0 | 24.4 | 5.0 |
| 2 | 0.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 9.3 | 6.0 | 70.8 | 0.0 | 0.0 | 13.9 | 0.4 |
| 4 | 1.3 | 0.0 | 0.0 | 98.7 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 0.6 | 0.6 | 0.0 | 98.8 | 0.0 | 0.0 |
| 6 | 40.2 | 0.0 | 7.1 | 0.0 | 0.0 | 52.8 | 0.0 |
| 7 | 0.0 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 | 73.0 |

In Table 3, 4 and 5, No. 1 to 7 represents the class number of the cover type as determined on the ground truth, and No. I - VII represents the class number of cover type as classified on the SPOT image.

HIGH RESOLUTION SATELLITE IMAGE TEXTURE FOR MODERATE RELIEF TERRAIN ANALYSIS

Derek R. Peddle

and

Steven E. Franklin

Remote Sensing Group
NORDCO Limited, P.O. Box 8833
St. John's, Newfoundland
Canada A1B 3T2

Department of Geography
The University of Calgary
Calgary, Alberta, Canada
T2N 1N4

ABSTRACT

Spatial co-occurrence matrices are computed for a SPOT HRV multispectral image of a moderate relief environment in eastern Canada. The texture features entropy and inverse difference moment are used with the spectral data in land-cover classification, and substantive increases in accuracy are noted. These range from 10% for exposed bedrock to over 40% in forest and wetland classes. The average classification accuracies are increased from 51.1% (spectral data alone) to 86.7% (spectral data plus entropy measured in band 2 and inverse difference moment in band 3). Classes which are homogeneous on the ground are characterized adequately by spectral tone alone; but classes which contain distinct vegetation patterns or are strongly related to structure are significantly improved using spectral texture.

Keywords: Spatial Co-occurrence, Entropy, Classification

INTRODUCTION

Spectral response patterns can be interpreted for textural features (e.g. Haralick *et al.*, 1973; Irons and Petersen, 1981; Gougeon and Wong, 1987) which are important in classification (e.g. Shih and Schowengerdt, 1983) and image transformations (e.g. Townshend, 1981). One commonly applied procedure uses image gray-level spatial dependency or co-occurrence. The image data are viewed as an instance of a random process and parameters of that process are estimated using summary statistics such as entropy derived from directional co-occurrence matrices. A program (Franklin and Peddle, 1987) to accomplish this has been written for application to a wide range of digital data including elevation models, aeromagnetic survey data, digitized aerial photography, synthetic aperture radar imagery, and satellite spectral response patterns. In each application, the rationale for using image texture is based on an understanding of the physical variable involved; for example, in processing a DEM for texture, a variable that approximates terrain relief or roughness is generated from elevation spatial co-occurrence.

In this paper we report on a study of SPOT HRV multispectral image texture for mapping and analysis of vegetation communities in Gros Morne National Park. An earlier study (Franklin and Peddle, 1989) showed the value of image texture in analysis of this area from Landsat MSS data. With SPOT imagery, our initial interpretation was that the spectral response would be more representative of true ground conditions and less a function of a composite of ground conditions (see for example, Connors *et al.*, 1987; Jones *et al.*, 1988). The high resolution image may contain within class variability that was suppressed

in the Landsat image, and also not mapped during aerial photointerpretation. Texture features may resolve potential class confusions and result in a higher mapping accuracy from SPOT imagery. Another way to consider this is as follows: High resolution image tone may be inherently more representative, but should be analysed with some reference to the context in which it occurs.

STUDY AREA AND METHODOLOGY

The study area in Gros Morne National Park is on the west coast of the island of Newfoundland in eastern Canada. The area contains boreal vegetation, but in the high elevations (above 600 metres a.s.l.) some tundra and northern species prevail. The maximum relief in the subarea selected for this study is approximately 50 metres. The SPOT image was acquired on 17 August 1988 with a viewing angle of 20.24 degrees west, solar elevation 52.9 degrees, solar azimuth 164.7 degrees, and is cloud-free.

A class structure derived primarily through aerial photointerpretation, field work and land systems analysis was used to drive the image analysis. The spatial co-occurrence matrices were analyzed for entropy and inverse difference moments in four orientations in all three SPOT bands using 3x3 and 5x5 windows; but these proved unstable, probably because a texture measure on this scale may be related only to the chance location of the window (see Townshend, 1981: p. 79). A 21x21 image window was used corresponding to 420x420 metres on the ground; this decision was based partly on our earlier success in mapping similar classes with texture from 3x3 and 5x5 MSS windows (Franklin and Peddle, 1989). Discriminant functions were generated for each class with (i) SPOT HRV data alone, (ii) texture alone, and (iii) SPOT HRV data plus entropy of band 2 in the right diagonal orientation and inverse difference moment of band 3 in the left diagonal orientation. Since no independent test areas were available in support of this analysis, training data were divided randomly into two equal sets of 100 pixels per class for training and testing.

RESULTS OF THE ANALYSIS

Table 1 contains a summary of the classification results derived using the various discriminant functions. The overall classification accuracy using the SPOT HRV data alone was 51.1%. The SPOT HRV entropy and IDM measures together were 48.9% correct and a combined spectral/textural data set was 86.7% correct. These results must be interpreted cautiously since it is known that classification accuracies derived in this

way will overestimate accuracy. However, the general trend is clearly defined and readily interpreted; overall, those classes that are homogeneous on the ground are accurately characterized with spectral tone alone, but those classes that are not spectrally distinct, that contain unique variability related to vegetation patterns or are related strongly to structural features such as topography, are improved significantly using image texture.

For example, class 2 (Heath) which is spatially heterogeneous is poorly defined using spectral data alone (32%). Texture alone shows 58% accuracy, and the combination of spectral and textural information increases this to 78%. Class 3 (Exposed Bedrock), on the other hand, illustrates a spectrally distinct class that is not textural. The spectral data alone yield 71% accuracy and texture features improve this only an additional 10%; texture alone results are just 18% correct. The Wetland class (class 7) shows the opposite effect where texture is a poor discriminator alone, but together with the spectral data shows a substantive overall increase. Finally, class 5 (Deciduous Forest) illustrates a textural class that is very accurate when analyzed with texture alone (95%) compared to spectral tone alone (44%) or the combined function (92%). This is a relatively small class that occurs in topographically unique conditions only at relatively high elevations in this part of the Park.

The final classification product is shown in Figure 1. This was produced using a maximum likelihood technique on the spectral and textural data on an ARIES-III system at NORDCO Limited (original plotted in colour) and, when compared to the results of aerial photointerpretation, the field data and the Park Biophysical Inventory, is an excellent approximation of the spatial arrangement and internal homogeneity of the classes.

CONCLUSION

Texture features of high resolution satellite imagery contain information that may be used to increase land-cover classification accuracy. In this paper, we illustrate the improvements possible when SPOT HRV multispectral imagery of a moderate relief, boreal terrain in eastern Canada are augmented with entropy and inverse difference moments derived from directional spatial co-occurrence matrices compiled over 21x21 pixel windows at original image resolution. The overall improvements are significant: from 51.1% accuracy to 86.7% accuracy for seven classes including the dominant vegetation communities. This study complements our earlier analysis of the impact of spectral texture on Landsat MSS classification of this area, and confirms the importance of image texture for separation and analysis of classes in this environment.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of Parks Canada, NORDCO Limited, the Natural Sciences and Engineering Research Council of Canada and the University of Calgary. SPOT imagery are copywrite (SPOT Corp., CNES, 1988).

REFERENCES

1. Connors, K. W., Gardner, T. W., and Petersen, G. W., 1987, Classification of geomorphic features and landscape stability in northwestern New Mexico using simulated SPOT imagery: *Remote Sensing Environ.*, 22: 187-207.
2. Franklin, S. E., and Peddle, D. R., 1987, Texture analysis of digital image data using spatial co-occurrence: *Computers & Geosciences*, 13(3): 293-311.
3. Franklin, S. E. and Peddle, D. R., 1989, Spectral texture for improved class discrimination in complex terrain: *Int. J. Remote Sensing (Remote Sensing Letters)*, in press.
4. Gougeon, F. and Wong, A., 1987, Spectral and textural segmentation of multispectral aerial images: *Proc., 10th Can. Symp. Remote Sensing*, Edmonton, 291-300.
5. Haralick, R. M., Shanmugam, K., and Dinstein, I., 1973, Texture features for image classification: *IEEE Trans. Systems, Man, Cybernetics.*, 3(2): 610-621.
6. Irons, J. R. and Petersen, G. W., 1981, Texture transforms of remote sensing data: *Remote Sensing Environ.*, 11(5): 359-370.
7. Jones, A. R., Settle, J. J., and Wyatt, B. K., 1988, Use of digital terrain data in the interpretation of SPOT-1 HRV multispectral imagery: *Int. J. Remote Sensing*, 9(4): 669-682.
8. Shih, E. and Schowengerdt, R. A., 1983, Classification of arid geomorphic surfaces using Landsat spectral and textural features: *Photog. Eng. Remote Sensing*, 50(1): 83-91.
9. Townshend, J. R. G., ed., 1981, *Terrain analysis and remote sensing*, George Allen & Unwin: London. 232p.

Table 1: Classification Accuracy by Class

| Function | 1 | 2 | 3 | 4 | 5 | 6 | 7 | X |
|------------------|----|----|----|----|----|----|----|------|
| SPOT Alone | 47 | 32 | 71 | 66 | 44 | 47 | 51 | 51.1 |
| Texture Alone | 38 | 58 | 18 | 74 | 95 | 50 | 9 | 48.9 |
| SPOT and Texture | 87 | 78 | 81 | 96 | 92 | 77 | 96 | 86.7 |

Note:

Classes 1 - 7 are described in figure 1

SPOT Alone = Bands 1, 2, and 3

Texture Alone = Entropy and Inverse Difference Moment

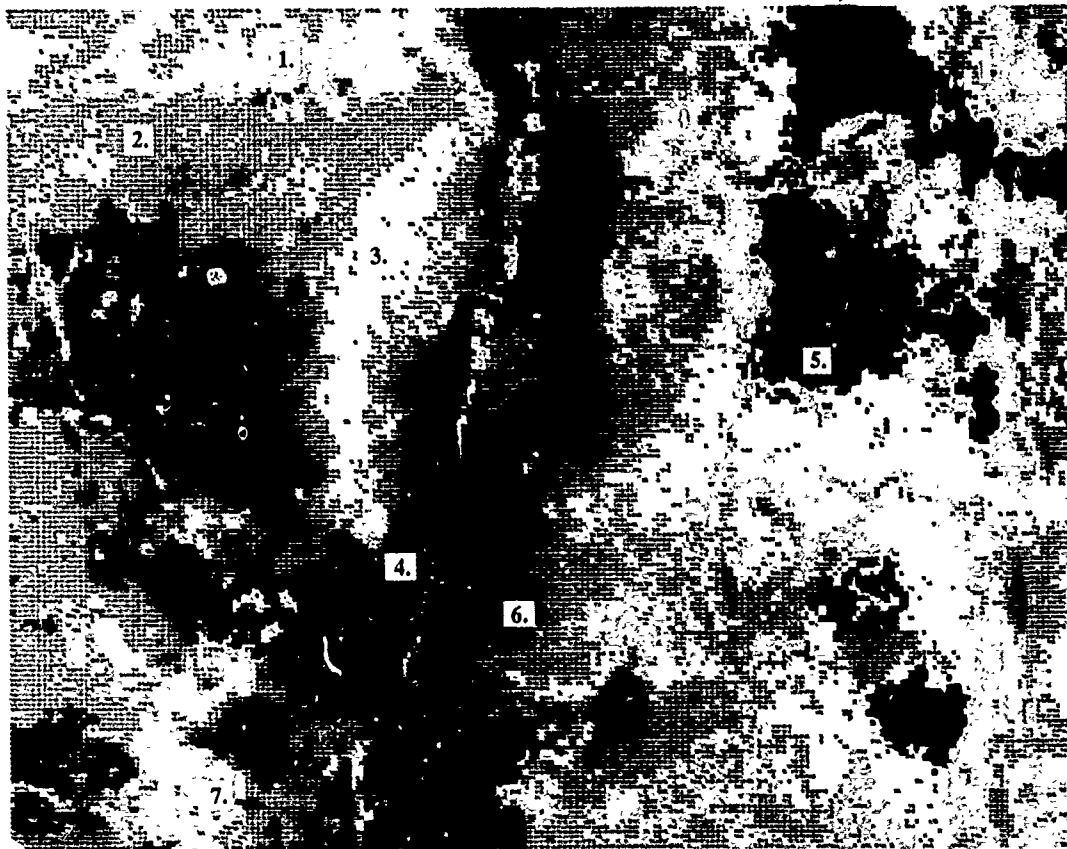
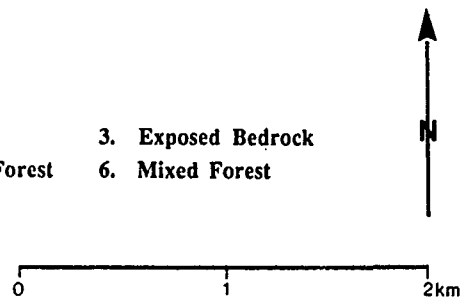


Figure 1: Supervised Classification

- | | | |
|------------------------|---------------------|--------------------|
| 1. Balsam Fir Forest | 2. Heath | 3. Exposed Bedrock |
| 4. Black Spruce Forest | 5. Deciduous Forest | 6. Mixed Forest |
| 7. Wetland | | |



INTEGRATION OF CONTEXT CLASSIFIERS WITH GIS

D.J.Booth, T.R.E.Chidley and W.G.Collins

Remote Sensing Unit
 Dept. of Civil Engineering
 Aston University
 Aston Triangle
 Birmingham
 B4 7ET
 United Kingdom

ABSTRACT

The paper discusses the various contextual reclassification algorithms which have been applied to remotely sensed data. These are applied to Landsat TM data, and compared in terms of their speed and their degree of enhancement. The possibilities of GIS for designing fast and efficient per-field classification algorithms are examined.

INTRODUCTION

One of the most important tasks of Geographical Information Systems and Remote Sensing is to produce working maps as output. Of necessity, these must be generalised from the classified output to be readily legible at the output scale. An added advantage of this generalisation is that, particularly in the case of remotely sensed data, the overall accuracy of the final product is often enhanced by the removal of mis-classified areas, caused by noisy data.

This paper looks at Contextual reclassification techniques based on classified remotely sensed data and suggests that the advent of digital maps and GIS could provide a simple and effective answer to the problem of designing a per-field classification algorithm, whilst avoiding the necessity of regionalisation algorithms.

CONTEXTUAL RECLASSIFICATION

Contextual reclassification is the process whereby a pixel's class is modified according to its neighbours. The simplest form of contextual reclassification is the use of a mode filter on a classified image (Rothery, 1982). In this case small areas tend to be replaced by the regionally dominant class, effectively removing the pixels which have been incorrectly classified due to noise and/or edge effects in the original image. This tendency to remove small areas results in some correctly classified pixels (those representing small areas on the ground) also being removed; a process akin to cartographic generalisation.

A similar, but more specifically targeted, contextual reclassification algorithm is that of small area replacement (Letts, 1979). In this method, contiguous areas containing less than a specified threshold number of pixels are 'declassified', and subsequently reclassified according to their neighbouring pixels (those still classified). This reclassification can be on the basis of 'nearest classified neighbour', or 'modal class of n nearest classified neighbours'.

A variation on this technique is to try to identify noise and boundary pixels on the original imagery. This is achieved by first passing a suitable edge detector over the image (eg Roberts, Sobell or Laplacian high pass filters), thresholding this to separate the image into edges/non-edges, declassifying the pixels identified as edges, then applying one of the above reclassification algorithms to the declassified image. The thresholding may be done by trial and error, or some automatic solution can be applied.

Another technique is that of Relaxation Labelling, where the image is reclassified according to the probabilities of a pixel and its neighbours belonging to each class (DiZenzo et al, 1987a, 1987b, Mohn et al, 1987, Smith et al, 1981, Kittler and Illingworth, 1985). This probabilistic relaxation is an extension of the maximum likelihood classification algorithm to cover a neighbourhood, rather than single pixels. Because the algorithm is extremely slow to implement for large neighbourhoods, attention has focussed on iterative implementation over small neighbourhoods, best results being achieved after several iterations.

To try to overcome the time problems caused by probabilistic relaxation, non-probabilistic methods have been investigated, for example modifying the minimum distance classifier in a similar manner to the maximum likelihood algorithm above, but classifying according to distances to means, rather than probabilities. Another alternative is to use ranked classes for each pixel, since rank order of classes can be calculated rather faster than the actual probabilities using an

equation derived from the ML probability estimate (Mather, 1985).

The image can be contextually reclassified using Markov chain theory to model the spatial autocorrelation of the class map (Kittler and Foglein, 1984). This is possible using the image itself, however, research suggests that it is more effective to estimate the transition probability matrix from field data, rather than the 'noisy' classified image. This method also has the added advantage of estimating the distance beyond which a pixel has no influence on its neighbours (the order of the chain). This distance could be used to set the size of mode filter window to be used, since the process of Markov relaxation is rather slow.

A contextual reclassifier, bearing some resemblance to textural classifiers, has been suggested by Wharton (1982). The so-called CONAN algorithm works by first creating a new 'frequency of occurrence' image, the bands of which have values according to each class' frequency of occurrence over an n by n window. The classifier is then 'retrained', using the same areas, on this new image, and the classification rerun on this. The result is an extremely efficient and relatively fast contextual reclassification.

The final contextual reclassification algorithm, and the most important in terms of integration with GIS, is the per-field classifier. This has long been considered desirable (Allan, 1986), but the non-trivial task of extracting regions from remotely sensed imagery has precluded its widespread use. The development of digital mapping and GIS now enables field boundaries to be extracted from maps, and classification to take place on a field by field basis. This classification can either be of a modal basis (the field's modal class is returned to every pixel within the field), or on a statistical basis (the overall most probable class is assigned to every pixel in the field).

METHODOLOGY

The algorithms used are all well documented in the literature cited, with the exception of Markov relaxation. In order to save space the reader is referred to these papers, and only a short explanation of the Markov relaxation procedure is given here:

Markov chains are a method of modelling processes where the next 'event' or 'state' is dependent on the previous event or state, hence the term chain. If 'state' is replaced by 'class', and 'next state' by 'adjacent class', the relevance to contextual reclassification is evident: the chain can be used to determine the probabilities of each class occurring adjacent to each other class.

These probabilities are represented by the Transition Probability Matrix, which can be derived by normalising a tally matrix with entries for sample pixels and their neighbours, an example of which is shown below:

Fig. 1. Tally Matrix

| | neighbour class | | | |
|---------------|-----------------|-----|-----|----|
| | 1 | 2 | 3 | 4 |
| pixel class 1 | 198 | 12 | 1 | 0 |
| 2 | 12 | 200 | 25 | 3 |
| 3 | 1 | 25 | 109 | 0 |
| 4 | 0 | 3 | 0 | 97 |

Multiplying the transition probability matrix by itself gives the probabilities of changes from one class to another after two 'moves', in other words the probabilities for pixels of one class lying two pixels away from pixels of another class. Continued multiplication gives probabilities for three pixels distance and so on. Eventually, at the limits of autocorrelation, the chain 'converges', and there is an equal probability of any class occurring at this distance from a pixel. This distance is the order of the chain.

Once these transition probabilities have been calculated it is a simple (although slow) procedure to examine all pixels within this distance (order of the chain) from a central pixel, sum the probabilities for each class and assign the central pixel to the class with the highest probability. This represents a distance decaying contextual reclassification algorithm, and initial experiments have been very promising.

EXPERIMENTAL PROCEDURE

Most of the above enhancements have been carried out on classified Landsat TM imagery of the UK Derbyshire Peak District (26.04.84), geometrically corrected to UK Ordnance Survey map grid with 25m square pixels for convenience. Ten cover classes were defined as broadly representative of the area. The authors would like to point out that the object of the experiment was to compare relative accuracies of the algorithms and that the data presented here should only be used for this purpose. For this reason, accuracies and timings are given relative to those for the lowest raw classification accuracy, thus an accuracy of 1.4 represents 1.4 times more accurate than the box classified image, and a time of 1.7, 1.7 times longer to produce the classification compared to this. All algorithms were run on a DEC vax 8650, timings being 'charged CPU time'. Accuracies were normalised from contingency tables drawn up using different data to that used to train the classifiers.

RESULTS

The table below shows the relative speeds and accuracies of the algorithms, applied to supervised box and maximum likelihood classifiers. The timings for the contextual enhancements are those for classifier plus enhancement.

Table 1. Results.

| Algorithm | Normalised Accuracy | Time |
|-----------------------|---------------------|-------|
| Box classifier | | |
| alone | 1.00 | 1.00 |
| 3*3 mode | 1.04 | 3.11 |
| 9*9 mode | 1.07 | 3.87 |
| SAR 4 | 1.32 | 1.54 |
| SAR 8 | 1.38 | 1.54 |
| Edge | 1.33 | 3.47 |
| CONAN | 1.28 | 4.97 |
| Markov (a) | 1.37 | 58.66 |
| Markov (b) | 1.37 | 58.66 |

Maximum Likelihood

| | | |
|------------|------|--------|
| alone | 1.39 | 48.94 |
| 3*3 mode | 1.43 | 51.05 |
| 9*9 mode | 1.45 | 51.80 |
| SAR 4 | 1.40 | 49.48 |
| SAR 8 | 1.40 | 49.48 |
| Edge | 1.39 | 51.41 |
| CONAN | 1.45 | 52.91 |
| Markov (a) | 1.14 | 106.60 |
| Markov (b) | 1.46 | 106.60 |

Explanation:

SAR 4: Small area replacement; areas containing less than 4 pixels reclassified by nearest neighbour.

SAR 8: Small area replacement; areas containing less than 8 pixels reclassified by nearest neighbour.

Edge: Roberts edge detector, automatically thresholded to identify areas to declassify. Reclassification by nearest neighbour.

CONAN: Wharton's CONAN algorithm, implemented over a 9 by 9 patch.

Markov (a): Markov relaxation. Transition probability matrices calculated from classified image.

Markov (b): Markov relaxation. Transition probability matrices calculated from field data.

DISCUSSION

The results should be taken as generally indicative of the sort of accuracy improvements to be expected from using these techniques, and the time penalties associated with them. It should be noted, however, that the authors are not professional programmers, and whilst the algorithms operate successfully, the attention of a professional would undoubtedly speed them up considerably. In addition to this, the continued increase in computer speed, combined with decreases in costs could well make these time constraints academic in future.

CONCLUSIONS

Contextual enhancement of classified remotely sensed data can considerably improve classification accuracy. Even simple techniques such as mode filtering can bring about this improvement. To some extent the improvements gained are limited by the accuracy of the data used in contextual enhancement (a good example being the

decrease in accuracy obtained when the Markov transition probability matrix was estimated from the image itself). The increases in accuracy obtained by the more complex methods are only slightly higher than those obtained by mode filtering and small area replacement. This difference may or may not be regarded as significant, depending on the work being undertaken by an operator.

Contextual enhancement also has the added advantage of generalising output for map production at any given scale. This will be of particular use in the field of GIS where such documents need to be produced on a routine basis. The advent of GIS may also enable practical per-field classifiers to be designed without the constraints of first having to perform non-trivial region extraction from imagery.

REFERENCES

Allan, J.A. (1986) 'How Few Data do we Need: Some Radical Thoughts on Renewable Natural Resources Surveys' Proc. Symposium on Remote Sensing for Resources Development and Environmental Management, Enschede.

Kittler, J. and Foglein, J. (1984) 'Contextual Classification of Multispectral Pixel Data' Image and Vision Computing, Vol.2 No.1.

Kittler, J. and Illingworth, J. (1985) 'Relaxation Labelling Algorithms-a Review' Image and Vision Computing, Vol.3 No.4.

Letts, P.J. (1979) 'Small Area Replacement in Digital Thematic Maps' Proc. 1979 Machine Processing of Remotely Sensed Data Symposium.

Mather, P.M. (1985) 'A Computationally-Efficient Maximum Likelihood Classifier Employing Prior Probabilities for Remotely Sensed Data' International Journal of Remote Sensing, Vol.6 No.2.

Mohn, E. Hjort, N.L. and Storvik, G.O. (1987) 'A Simulation Study of Some Contextual Classification Methods for Remotely Sensed Data' IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25 No.6.

Rothery, D.A. (1982) 'Supervised Maximum-Likelihood Classification and Post-Classification Filtering Using MSS Imagery for Lithological Mapping in the Oman Ophiolite' Proc. International Symposium on Remote Sensing of Environment, Second Thematic Conference, Remote Sensing for Exploration Geology, Fort Worth, Texas.

Smith, B.W. Siegel, H.J. and Swain, P.H. (1981) 'Contextual Classification on a CDC Flexible Processor System' Proc. 1981 Machine Processing of Remotely Sensed Data Symposium.

Wharton, S.W. (1982) 'A Context-Based Land-Use Classification Algorithm for High-Resolution Remotely Sensed Data' Journal of Applied Photographic Engineering, Vol.8 No.1.

Dizenzo, S. Bernstein, R. Degloria, S.D. and Kolsky, H.G. (1987a) 'Gaussian Maximum Likelihood and Contextual Algorithms for Multicrop Classification' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.6

Dizenzo, S. Bernstein, R. Degloria, S.D. and Kolsky, H.G. (1987b) 'Gaussian Maximum Likelihood and Contextual Algorithms for Multicrop Classification Experiments Using Thematic Mapper and Multispectral Scanner Sensor Data' IEEE Transactions on Geoscience and Remote Sensing, Vol.GE-25 No.

CHANGES IN CANOPY AND BRANCH REFLECTANCE ALONG A SO₂ GRADIENT ON NORTHERN VANCOUVER ISLAND

Submitted for presentation at IGARSS '89 by:

Brian Curtiss¹ and Susan L. Ustin²

¹Center for Study of Earth from Space (CSES)
Cooperative Institute for Research in the Environmental Sciences (CIRES)
University of Colorado Boulder, CO 80309
(303) 492-6718 (303) 492-5070 (FAX)

²Department of Botany
University of California Davis, CA 95616
(916) 752-0621

Understanding the effects of air pollutants on natural ecosystems and predicting the possible consequences of various levels of exposure is of major biological concern. The area in the vicinity of the Sulphite pulp mill at Port Alice, British Columbia provides an ideal setting for the examination of the effects of SO₂ on coniferous forests. Gaseous sulfur dioxide, when taken up by the needles, results in the formation of bisulfite, a highly toxic compound that effects the functioning of the cell wall and other cellular membranes. The pulp mill at Port Alice, which has been active since 1918, is located on the eastern shore of Neroutsos Inlet on the northern part of Vancouver Island. The forests in the immediate vicinity of the mill show symptoms of chronic exposure, including chlorosis, a reduction in photosynthetic biomass, mortality and a decrease in species diversity. While the effects of SO₂ exposure on individual trees is readily apparent in the immediate vicinity of the mill, the effect on the surrounding forests is difficult to quantify using conventional forest inventory methods. Reflectance spectra and chlorophyll a and b measurements were made at five sites, composed predominantly of western hemlock, along the SO₂ exposure gradient. The variance of all other site variables was minimized. Trends were observed in the reflectance spectra that are attributable to a decreased chlorophyll content and an increase in chloroplast damage at those sites nearer to the pulp mill. The decrease in total chlorophyll concentration resulted in a decrease in the depth of the chlorophyll absorption well. Chloroplast damage resulted in a broadening of the chlorophyll absorption feature. This broadening is commonly associated with increasing disorder of the chloroplast membranes and is a result of damage associated with exposure of the membranes to bisulfite. Thus, it is shown that data acquired using a high spectral resolution sensor could be utilized to assess and map the extent of forest damage.

IGARSS'89

LABORATORY FLUORESCENCE, REFLECTANCE, AND CHLOROPHYLL MEASUREMENTS
OF FOLIAGE FROM A STRESSED SPRUCE FOREST

C. Banninger

Institute for Image Processing and Computer Graphics
Wastiangasse 6, A-8010, Graz, Austria
Tel. (316) 8021, Telex (3)11265, Telefax (316) 8021-20

Chlorophyll pigments are the means by which a plant uses solar radiation to convert carbon dioxide and water into biochemical compounds that are essential to its health and vitality. A dysfunctioning of the photosynthetic process in response to environmentally related physiological changes in a plant generally result in a reduction in chlorophyll pigment content and an associated increase in plant reflectance and fluorescence in the visible to near-infrared wavelength region. The greatest spectral changes occur at wavelengths that correspond to maximum pigment absorptance and fluorescence, and their measurement could be used to assess the degree of physiological damage in a plant and the level of environmental stress present. This would establish an analytical basis for interpreting similar data acquired by air- and spaceborne sensor systems from vegetation canopies.

Laboratory reflectance and fluorescence measurements were made of foliage samples collected from a metal-stressed Norway spruce forest and their chlorophyll-a, -b, and carotenoid contents determined. Forest soil samples were also analysed as to their mineral content. Reflectance measurements were made over the 400 - 900 nm wavelength region at a spectral sampling interval of 2 nm and over the 400 - 800 nm region with 10 ns excitation pulse widths between 337 and 700 nanometres for the fluorescence measurements.

Correlations are strongest between needle pigment content and fluorescence values at 740 nm and reflectance values at around 675 nm and 760 nm, but are much less strong at 440 nm and 525 nm for the former and 550 nm for the latter. Strong correlations were also obtained between needle pigment contents and soil mineral content. Changes in the position of the reflectance red edge were ambiguous, which is probably due to the stability of the chlorophyll-a to chlorophyll-b ratio between stressed and non-stressed needles. The potential to measure concentrations and changes in canopy pigment (in particular chlorophyll) concentrations by high resolution passive and active optical sensors is good, but the use of the reflectance red edge as a indicator or measure of canopy stress appears doubtful.

Reflectance and Transmittance of Snow at High Spectral Resolution

Jeff Dozier

Robert E. Davis

Anne W. Nolin

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

University of California
Sierra Nevada Aquatic Research Laboratory
Mammoth Lakes, CA 93546

University of California
Center for Remote Sensing and Environmental Optics
Santa Barbara, CA 93106

ABSTRACT

A radiative transfer model for the bidirectional reflectance-distribution function (BRDF) shows that snow is moderately anisotropic in the near-infrared wavelengths. Although the directional-hemispherical albedo of snow decreases as the grains become larger, the forward scattering also increases, with the result that the illumination and viewing geometry must be considered when interpreting physical properties of the surface layer of the snow pack from remote sensing data. Measurements of the BRDF and the transmittance for a variety of snow conditions were made throughout the winter and spring seasons with a SE-590 spectro-radiometer, for wavelengths from 0.38 to 1.11 μm . Coincident with these the surface grain properties were analyzed by stereological methods. The sphere with the same surface-to-volume ratio as the ice grains is used as the equivalent sphere.

INTRODUCTION

Our interest in the spectral reflectance and transmittance of snow, and their relationship to the snow's physical properties, stems from two reasons:

1. We need to estimate the albedo of snow from satellite in order to estimate the net solar radiation flux at the surface, given by the convolution of the spectral distribution of the incoming radiation and the spectral albedo. The spectral albedo for direct and diffuse irradiance can be modeled from a knowledge of the snow's physical properties [1]. However, from satellite or aircraft we can measure only a part of the spectral BRDF (bidirectional reflectance-distribution function), at a few wavelengths and usually only one solar-viewing orientation. We therefore need to know the spectral BRDF of snow as a function of its physical properties. We can use this BRDF as a boundary condition in an atmospheric radiation model, to predict the at-satellite radiance as a function of surface snow properties and the atmospheric profile.

2. Unlike soil surfaces, solar radiation in the visible wavelengths can penetrate tens of centimeters into the snowpack. Infrared radiation, however, is emitted only from the top millimeter or so, so infrared cooling occurs only from the surface. This asymmetric heating and cooling process causes steep temperature gradients near the snow surface at night, and therefore much different metamorphic rates than would occur if the heating and cooling were symmetric. Moreover, the penetration of solar radiation can cause photochemical reactions with such contaminants as H_2O_2 , and evaluation of these rates is important for interpretation of ice-core data.

OPTICAL PROPERTIES OF SNOW AND ICE FROM 0.4 TO 2.5 μm

In visible wavelengths ice is one of the most transparent substances in nature. For example, the e -folding distance, at which transmittance through pure ice is reduced to e^{-1} , for wavelength $\lambda = 0.46 \mu\text{m}$ is 24 m. Snow reflectance in the visible wavelengths is therefore sensitive to contaminants and to shallow depth and insensitive to grain size.

Through the near-infrared wavelengths ice is moderately to highly absorptive. The e -folding distance is 1.4 cm at $\lambda = 1.2 \mu\text{m}$, 0.34 mm at $\lambda = 1.6 \mu\text{m}$. Snow reflectance in the near-infrared is therefore sensitive to grain size and insensitive to contaminants.

We can therefore use data at appropriate spectral wavelength and resolution to measure grain size and contaminants in the surface layer. The reflectance of snow is modeled as a multiple scattering problem, as first proposed by Bohren and Barkstrom [2] and reviewed in detail by Warren [1]. The calculations of hemispherically integrated albedo match measured spectra [3], but the reflectance measurements were seldom accompanied by rigorous measurements of snow properties.

The scattering properties for an ice sphere of appropriate radius r can be calculated by the Mie equations [4], the complex angular momentum approximation [5], or, for larger grains, by geometric optics [6]. Then the radiative transfer equation [7] can be used to calculate the multiple scattering and absorption of the incident radiation.

$$\mu \frac{dL(\tau, \mu, \phi)}{d\tau} = -L(\tau, \mu, \phi) + J(\tau, \mu, \phi) \quad (1)$$

L is radiance at optical depth τ in direction θ, ϕ ; ϕ is the azimuth; θ is the angle from zenith, and $\mu = \cos\theta$. J is the source function; it results from scattering of both direct and diffuse radiation or, at thermal wavelengths, emission.

The presence of liquid water in the snow should not by itself affect the reflectance. Except where meltwater ponds in depressions when melting snow overlies an impermeable substrate, liquid water content in snow rarely exceeds 5 or 6%. This small amount of water does not appreciably affect the bulk radiative transfer properties, except possibly in those wavelength regions where the absorption coefficients are appreciably different. Instead, the changes in reflectance that occur in melting snow result from the increased crystal sizes and from an effective size increase caused by the two- to four-grain clusters that form in wet, unsaturated snow [8,9]. These apparently behave optically as single grains, causing decreased reflectance in near-infrared wavelengths. O'Brien and Munis [3] observed the spectral reflectance of a snow sample to be lower after warm air had been blown over it, but that the reflectance did not increase when the snow was refrozen.

There is no explicit dependence on density, at least for the semi-infinite snowpack. Although Bergen [10] has proposed a semi-empirical formulation with a density term for snow reflectance, in practice the natural increases in density are usually accompanied by increases in grain size.

In our examination of the spectral properties of snow, it is computationally time-consuming to calculate the angular distribution of the reflected radiation. But it is comparatively simple to examine the reflectance integrated over all angles, and these calculations show the interesting spectral properties. We therefore present two types of calculations:

1. Those that show the angular dependence result from a discrete-ordinates solution to the radiative transfer equation [11]. The quantity calculated is the spectral bidirectional reflectance distribution-function (BRDF):

$$f_r, \lambda(\mu, \phi) = \frac{L_\lambda(\mu, \phi)}{\mu_0 S_\lambda} \quad (2)$$

S_λ is the parallel beam at the top of the snowpack, incident at angle $\mu_0 = \cos\theta_0$.

2. Calculations of fluxes only, which show the spectral features but not the angular ones, result from a two-stream solution to the radiative transfer equation [12]. The quantity calculated is the spectral directional-hemispherical reflectance [13]:

$$R_{\tau, \lambda}(\mu_0) = \frac{\int_0^{2\pi} \int_0^1 \mu L(\mu, \phi) d\mu d\phi}{\mu_0 S_\lambda} = \frac{F_{\downarrow, \lambda}}{\mu_0 S_\lambda} \quad (3)$$

The transmittance is defined similarly:

$$t_{\tau, \lambda}(\mu_0) = \frac{F_{\downarrow, \lambda} + \mu_0 S_\lambda e^{-\tau(\lambda)\mu_0}}{\mu_0 S_\lambda} \quad (4)$$

The optical thickness of a snow layer, $\Delta\tau$, is a function of the extinction efficiency and the snow water equivalence W . The snow water equivalence W is the product of mean snow density, ρ_{snow} , and depth d . For the optical depth to be dimensionless, W is in kg m^{-2} and r is in meters.

$$\Delta\tau = \frac{3W Q_{\text{ext}}}{4r \rho_{\text{ice}}} = \frac{3\rho_{\text{snow}} d Q_{\text{ext}}}{4r \rho_{\text{ice}}} \quad (5)$$

How thick is a "semi-infinite" snow pack, for which the underlying surface has no effect? For practical purposes, we define it as a snow pack whose directional-hemispherical reflectance is within 1% of that at $\Sigma\Delta\tau = \infty$. For a solar zenith angle of 60° , Table 1 shows the minimum values in millimeters of water equivalence.

Table 1. Snow-Water Equivalence (mm) of Semi-Infinite Snow Pack

| λ (μm) | grain radius (μm) | | |
|--------------------------------|--------------------------------|-----|------|
| | 50 | 300 | 1000 |
| 0.45 | 17 | 63 | 145 |
| 0.7 | 10 | 37 | 80 |
| 0.9 | 5 | 15 | 30 |
| 1.6 | <1 | <1 | 1 |

Figure 1 shows the spectral reflectance of pure, deep snow for visible and near-infrared wavelengths at illumination angles 30° and 60° , for snow grain radii from 50 to 1,000 μm , representing the range for the finest new snow to coarse spring snow. Because ice is so transparent in the visible wavelengths, increasing the grain size does not appreciably affect the reflectance. The probability that a photon will be absorbed, once it enters an ice grain, is small, and that probability is not increased very much if the ice grain is larger. In the near-infrared, however, ice is moderately absorptive. Therefore, the reflectance is sensitive to grain size, and the sensitivity is greatest at 1.0 to 1.3 μm .

TRANSMITTANCE

Solar irradiance is a major contributor to the onset of snowmelt, particularly in alpine watersheds. Since snow is a weakly absorbing, strongly forward scattering medium, shortwave energy can penetrate to significant depths in the snowpack. The incident photons that do not exit from the snowpack will be absorbed in ice grains and converted to heat. This radiative heating can be responsible for significant changes in a snowpack over time. Because the effective thermal diffusivity of snow is low, this heat energy will be retained at depth in the snowpack. There is an asymmetry between radiative heating and cooling in the snow profile. If sufficient radiative heating occurs, it is possible to have melt occurring although surface air temperatures may not exceed 0°C .

When the snowpack transmits more radiant energy to greater depths, there is less chance of the photons being scattered back out of the snow surface, so the rate of radiative heating will be higher at depths where the net flux is highest. Ice is strongly absorbing in the infrared wavelengths, and these wavelengths are absorbed near the surface. The dual process of radiative heating by transmission and absorption of visible wavelengths in the snowpack and radiative cooling by emission from the snow surface will induce a vertical thermal gradient in the snowpack. This has important consequences for snow grain metamorphism such as surface hoar formation and ice grain sintering, both of which occur more rapidly in the presence of a thermal gradient [14, 15]. Formation of surface hoar also requires a steep thermal gradient in the upper portion of the snow profile and is enhanced by radiative heating. Snowpack optical properties are influenced by this heating process because metamorphism increases grain size, which in turn decreases reflectance in the near-infrared portion of the spectrum [16].

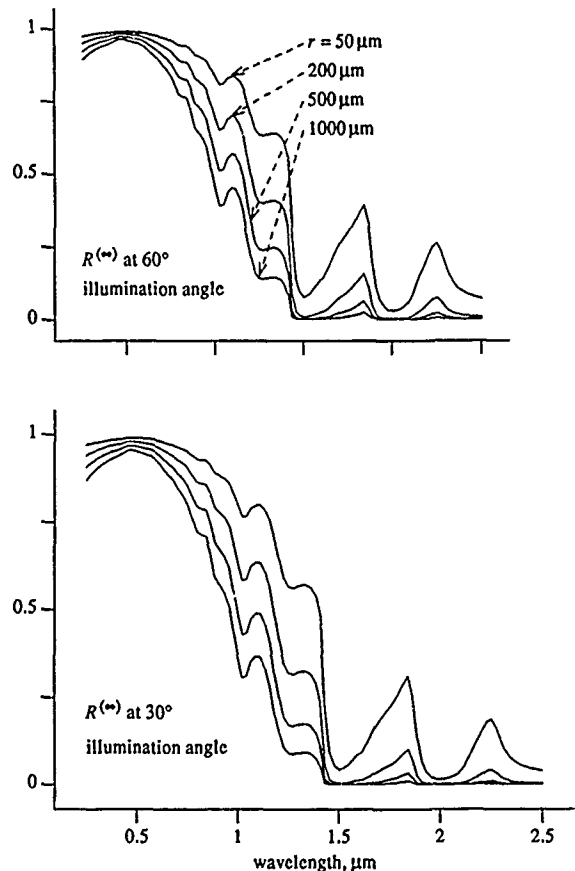


Figure 1. Spectral Reflectance of Deep Snow

Transmission of light into the snowpack is of interest for reasons other than hydrologic and optical. Snow chemistry is influenced by heating and grain metamorphism, as ions in the ice grains move outwards, concentrating on ice-grain surfaces. Biological responses to changes in the light regime beneath a thin snowpack would be important particularly in tundra landscapes.

Thermal measurements in a snowpack in Antarctica by Schlatter [17] showed highest temperatures at about 10 to 20 cm below the snow surface, indicating that radiative heating is an important process in such clean, dry, low-density snowpacks. Model calculations have shown that infrared radiative heating of snow and ice on the Martian surface could have been responsible for generation of liquid water while surface temperatures remained below freezing [18]. Transmittance and radiative heating are also important in creating a solid-state greenhouse in icy regoliths of Jupiter's satellite, Europa [19].

Using a multi-layer two-stream radiative transfer model, we calculated the net flux at levels within two different layered snowpacks. The spectral net flux at a level τ in the snow profile is:

$$F_{\text{net}, \lambda} = \frac{F_{\downarrow, \lambda}(\tau) + \mu_0 S_\lambda e^{-\tau(\lambda)\mu_0} - F_{\uparrow, \lambda}(\tau)}{\mu_0 S_\lambda} \quad (6)$$

F_{\downarrow} and F_{\uparrow} are hemispherically integrated upward and downward fluxes. Layer characteristics are given in Table 2, and the net flux profile for each snowpack is shown in Figure 2.

Net flux is a spectrally dependent quantity which has been calculated by summing over visible and infrared wavelengths, from 0.4 to 2.5 μm . Model results show that high-density layers increase absorption, and this effect is strongest at the surface.

Table 2. Properties of Two Snowpacks in Simulation

| Snowpack | Depth (cm) | Density (kg/m ³) | Grain Radius (mm) |
|-----------------|------------|------------------------------|-------------------|
| A (6-layers) | 0 - 3 | 450 | 0.5 |
| | 4 - 8 | 350 | 0.4 |
| | 9 - 18 | 350 | 0.35 |
| | 19 - 48 | 350 | 0.3 |
| | 49 - 78 | 250 | 0.3 |
| B (3-layers) | 79 - 98 | 200 | 0.3 |
| | 0 - 30 | 250 | 0.25 |
| | 31 - 70 | 400 | 0.4 |
| | 71 - 100 | 250 | 0.5 |

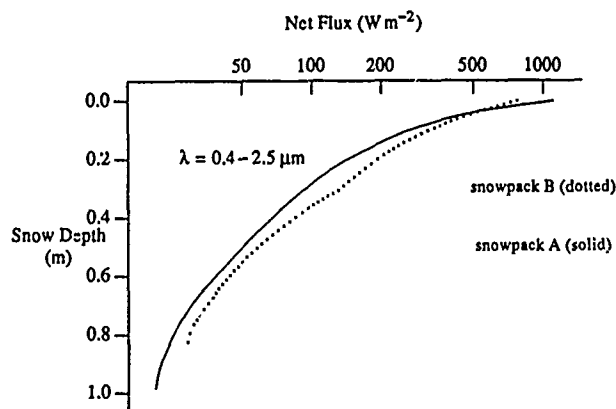


Figure 2. Net Flux (summed over wavelength) vs. Depth

Snowpack A (six layers) has a high-density layer at the top 3 cm, and this causes strong absorption near the surface. Net flux decreases more slowly with depth for Snowpack B (three layers), which has a low-density layer at the surface (0 - 30 cm). Large density differences between adjacent snow layers deeper within the snowpack will also affect light transmission. Net flux drops off quickly in Snowpack B as it reaches the higher density middle layer. Effects on net flux due to layer density variations are less apparent with Snowpack A since the flux has been so strongly attenuated by the top layer. Ice lenses and buried melt-freeze crusts would strongly attenuate the transmission of radiant energy and would have a higher rate of radiative heating than would a low-density, fine-grained layer in the snowpack. While albedo is not influenced by density differences, this is an important physical property of the snow for transmittance calculations and measurement.

Using a high spectral resolution radiometer, measurements were made of spectral net transmittance through snow for the visible and near-infrared wavelengths. These measurements compared well with calculated values, which showed peak transmission in the blue wavelengths where ice is most transparent, followed by a rapid decline in transmittance in the red and near-infrared wavelengths as ice becomes more strongly absorbing. The scattering properties of the snowpack were modeled by determining the optically equivalent sphere with the same surface-to-volume ratio, as measured by stereology [20].

For the visible wavelengths, solar irradiance penetrates the snowpack to depths greater than 10 cm. It is this shortwave portion of the incident irradiance which is responsible for radiative heating within the snowpack. Longer wavelengths are absorbed just below the snow surface and this heat exits the snowpack as longwave emitted energy, mostly at night thus, it is not considered a major cause of radiative heating of the snowpack. Calculations show that a high density layer at the surface causes a steep decline in net transmittance with increasing depth as compared to a low-density surface layer.

From these modeled cases of net transmittance for snowpack profiles, it is clear that interlayer variations induce radiative heating differences and, over time, would lead to increased stratification. This would be especially prevalent when absorbing impurities have accumulated on the surface of each layer, which subsequently become buried under new snowfall, since absorbing impurities will decrease the transmitted flux and increase absorption [1].

REFERENCES

- [1] S. G. Warren, "Optical properties of snow," *Rev. Geophys. Space Phys.*, vol. 20, no. 1, pp. 67-89, 1982.
- [2] C. F. Bohren and B. R. Barkstrom, "Theory of the optical properties of snow," *J. Geophys. Res.*, vol. 79, no. 30, pp. 4527-4535, 1974.
- [3] H. W. O'Brien and R. W. Munis, "Red and near-infrared spectral reflectance of snow," Research Rep. 332, 18 pp., Hanover, NH: U.S. Army Cold Regions Res. Engrg. Lab., 1975.
- [4] W. J. Wiscombe, "Improved Mie scattering algorithms," *Appl. Optics*, vol. 19, no. 9, pp. 1505-1509, 1980.
- [5] H. M. Nussenzveig and W. J. Wiscombe, "Efficiency factors in Mie scattering," *Phys. Rev. Lett.*, vol. 45, no. 18, pp. 1490-1494, 1980.
- [6] C. F. Bohren, "Multiple scattering of light and some of its observable consequences," *Amer. J. Phys.*, vol. 55, no. 6, pp. 524-533, 1987.
- [7] S. Chandrasekhar, *Radiative Transfer*, 393 pp., New York: Dover, 1960.
- [8] S. C. Colbeck, "Grain clusters in wet snow," *J. Colloid Interface Sci.*, vol. 72, no. 3, pp. 371-384, 1979.
- [9] S. C. Colbeck, "Classification of seasonal snow cover crystals," *Water Resour. Res.*, vol. 22, no. 9, pp. 595-705, 1986.
- [10] J. D. Bergen, "Observations on the relation of the short-wave reflectivity of recently deposited snow to its physical properties," *J. Climate Appl. Meteorol.*, vol. 22, pp. 193-200, 1983.
- [11] K. Stamnes, S. C. Tsay, W. Wiscombe, and K. Jayaweera, "A numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media," *Appl. Optics*, 1988.
- [12] W. E. Meador and W. R. Weaver, "Two-stream approximations to radiative transfer in planetary atmospheres: a unified description of existing methods and a new improvement," *J. Atmos. Sci.*, vol. 37, no. 3, pp. 630-643, 1980.
- [13] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance*, Natl. Bureau Stds. Monogr. 160, 52 pp., Washington, DC: U.S. Dept. Comm., 1977.
- [14] J. C. Giddings and E. R. LaChapelle, "The formation rate of depth hoar," *J. Geophys. Res.*, vol. 67, no. 6, pp. 2377-2383, 1962.
- [15] S. C. Colbeck, "On the micrometeorology of surface hoar growth on snow in mountainous area," *Boundary-Layer Meteorol.*, vol. 44, pp. 1-12, 1988.
- [16] J. Dozier, "Snow reflectance from Landsat-4 Thematic Mapper," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-22, no. 3, pp. 323-328, 1984.
- [17] T. W. Schlatter, "The local surface energy balance and subsurface temperature regime in Antarctica," *J. Appl. Meteorol.*, vol. 11, pp. 1048-1062, 1972.
- [18] G. D. Clow, "Generation of liquid water on Mars through the melting of a dusty snowpack," *Icarus*, vol. 72, pp. 95-127, 1987.
- [19] R. H. Brown and D. L. Matson, "Thermal effects of insolation propagation into the regoliths of airless bodies," *Icarus*, vol. 72, pp. 84-94, 1987.
- [20] R. E. Davis and J. Dozier, "Stereological characterization of dry alpine snow for microwave remote sensing," *Adv. Space Res.*, 1989. in press

REFLECTANCE SPECTRA OF VEGETATION GROWING ON MINE SITES IN THE CANADIAN SHIELD

V. Singhroy and F. Kenny

Ontario Centre for Remote Sensing
 Ministry of Natural Resources
 90 Sheppard Avenue East, 4th Floor
 North York, Ontario, Canada
 M2N 3A1

and

J. Springer

Ontario Geological Survey
 Ministry of Northern Development and Mines
 200 Elm St., Sudbury, Ontario, Canada
 P3C 5N3

ABSTRACT

Several studies have shown that vegetation which has been geochemically stressed may respond with changes in spectral reflectance. These changes have been observed at the green reflectance peak (570 nm), the chlorophyll absorption maximum (680 nm), the red reflectance edge (680 - 750 nm) and the infrared reflectance shoulder (750 - 1100 nm).

To determine the effects of different type and levels of mineralization on vegetation spectral response, laboratory measurements were conducted on vegetation samples from mineralized and non-mineralized sites within the Canadian Shield. These sites included the White Lake Mine (Zn, Cu, Pb), the Atikokan Fe Mine, the Whistle Mine (Ni, Cu), the Arnprior Mine (Zn, Pb) and the Natal mineralized site (Zn, Cr, Mn, Co, Cu, Pb, Ni, Au). Wavelength parameters, at 570 nm, 680 nm, 800 nm and at the point of inflection of the red edge, were calculated from 350 spectral curves measured from both mineralized and background sites.

Results for the White Lake area show extensive changes in reflectance spectra of the vegetation on the mineralized site, changes pronounced enough to permit the remote detection of the geochemical stress. On the other sites, the differences in reflectance spectra between the vegetation on mineralized and background areas were too slight to be detected by airborne sensors. These findings suggest that some surface mineralization within the Canadian Shield does not produce geochemical stress in vegetation to the extent that remote detection is feasible.

KEYWORDS: spectral reflectance, MEIS II, geobotany, mineral exploration, Canadian Shield

INTRODUCTION

Careful measurements of the spectral reflectance of leaves are required to document the changes in optical properties caused by geochemical stress. These changes have been observed at the green reflectance peak (570 nm), the chlorophyll absorption maximum (680 nm), the red reflectance edge (680-750 nm), and the infrared reflectance shoulder (750-1100) (Singhroy *et al.*, 1986, 1988; Horler *et al.*, 1980, 1983; Rock *et al.*, 1988; Collins *et al.*, 1983). These findings provide an understanding of the interaction of radiation with foliage for effective use of the data from high-resolution sensors, such as MEIS II, FLI and AIS. In this study, laboratory spectral reflectance measurements were made to establish the characteristic spectral response of geochemically-stressed vegetation at several abandoned mine sites, as a prerequisite for the remote detection of that response.

METHODOLOGY

The five mineralized sites reported in this paper (Figure 1) represent different bedrock geological settings and concentrations of heavy metals but are relatively constant in the Quaternary geology setting. In all test sites the surficial geology is represented by a relatively thin veneer of locally-derived basal till resting unconformably on a mineralized bedrock. Geochemical samples of bedrock and till on both background and mineralized sites were collected on a 10-metre grid system. Laboratory analysis of the base metal concentrations in these samples was subsequently performed.

Vegetation sampling was conducted in late August, just prior to fall senescence (Schwallier and Tkack, 1980 and Labovitz et al., 1983). Leaves were sampled from the most abundant tree species at each site. These included jack pine (*Pinus banksiana* Lamb), white birch (*Betula papyrifera*), trembling aspen (*Populus tremuloides*) and white pine (*Pinus strobus* L.).

Laboratory spectral reflectance measurements involved the use of both the Barringer Hand Held Ratioing Radiometer (HHRR) (Gladwell et al., 1983) and the Spectron SE 590 spectroradiometer. The HHRR uses a series of filters and records reflectance values in 3-nm-wide bands centered at 650, 680, 703, 713, 743, 782, and 800 nm. The SE 590 records reflectance values every 2.8 nm over a 400 to 1100 nm range. Both systems were mounted vertically three feet over the targets to measure a 6 cm by 6 cm area, and the light source was oriented 15 degrees off nadir to avoid shadows and variable illumination on the sample. The leaves were placed facing upward so that the entire 30 cm by 15 cm black aluminum plate of the radiometer was covered. Three layers or more of deciduous leaves were stacked to cover the whole plate. In the case of the conifers, the twigs were stacked as high as possible, usually four or five layers. This stacking was done to maintain the highest reflectance from the leaves in the near infrared portion of the spectrum (Hoffer 1978).

SPECTRAL REFLECTANCE MEASUREMENTS AND RESULTS

White Lake Site

The White Lake area is located in the Flin-Flon Volcanic Belt of Northern Manitoba and within the Amisk group of mafic intrusive pyroclastic and tuffaceous volcanic rocks (Bailes, 1971). The test site itself consists of a northerly-trending sulphide and chert-bearing argillite bounded by a fine to medium-grained gabbro sill. The heavy mineral concentrations on the mineralized site are in excess of 400 ppm Cu, 2000 ppm Zn, and 100 ppm Pb; whereas, the background site contained concentrations of < 200 ppm Cu, < 100 ppm Zn, and < 10 ppm Pb.

Both laboratory reflectance measurements and an airborne multispectral survey were conducted at this test site. The reflectance measurements were completed using the Barringer HHRR. From the four tree species sampled, the results showed significant differences in reflectance between the background and mineralized sites. These changes were observed at the chlorophyll absorption zone (680 nm), in the apparent position of the red edge (680 to 733 nm), and in magnitude changes at the infrared shoulder reflectance (780-800 nm). Figure 2 shows average reflectance spectra for jack pine needles along with standard error bars (confidence interval = 95%) and significant bandpass for remote detection of geochemically stressed vegetation.

Airborne narrow-band (10-nm), high-resolution (5-m) MEIS II imagery (Till et al., 1984), processed by combining band ratios and principal component techniques, were successfully used to detect the geochemically-stressed vegetation at the White Lake site (Figure 24).

Atikokan Iron Mine

The Atikokan site (Figure 1) is situated within the Quetico Greenstone Belt. Geochemical values of basal till in this area have concentrations of over 5% Fe, 1% Mg, 500 ppm Mn, and 100 ppm Zn.

Laboratory spectral reflectance measurements of sampled leaves were made with a SE 590 spectroradiometer. The results, summarized in Table 1, showed only minor differences in spectral reflectance between samples from the background and the mineralized sites (Figures 3 and 4). As the spectral reflectance measurements were so subtle, no airborne multispectral data was flown for this site.

Arnprior Site (Kingdom Mine)

The Kingdom Mine was a former lead, zinc and barium producer. The ore was found in Mississippi-Valley-type, open-space-filling veins, which had cuts through both the Precambrian basement and overlying Paleozoic rocks. Mineralization found in these veins included sphalerite, galena, fluorite, barite, and pyrite. Average geochemical values over the mineralized test site consisted of 106 ppm Zn, 1,490 ppm Pb, 130 ppm Ba and 190 ppm Fl. The background site had values of 11 ppm Zn, < 10 ppm Pb, and 130 ppm Ba.

The laboratory spectral reflectance measurements showed no significant shifts in the red edge (Table 1). White birch and trembling aspens, however, showed marked increases in the infrared plateau. Enhancements of airborne multispectral data (MEIS II and FLI) acquired over this site confirmed the higher reflectance readings from trembling aspen and white birch.

Whistle Mine

The Whistle Mine is a massive sulphide deposit at the northeastern rim of the Sudbury basin. It is a producing open-pit mine. Till and rock samples were taken immediately above the ore zone.

Spectral reflectance measurements from white pine and white birch showed minor evidence of changes in reflectance between samples taken from background and mineralized sites (Table 1 and Figures 3 and 4). As a result, no airborne flights were conducted.

Natal Lake

The Natal Lake site (Figure 1), located in the southwestern part of the Abitibi metavolcanic-metasedimentary belt, had a high concentration of sulfide minerals. The Precambrian geology, as described by Carter (1976), consists of a mafic to felsic metavolcanic suite together with ultramafic and minor intrusive felsic rocks, intruded by diabase dikes.

Geochemical assays for the mineralized area produced the following results: > 300 ppm Zn, > 400 ppm Cr and > 500 ppm Ni. At the background sites, the assay values were in the range of < 100 ppm Ni, < 200 ppm Cr and < 110 ppm Zn.

Results of the SE 590 laboratory measurements show no evidence of geochemical stress in the coniferous species at the Natal Lake site (Table I and Figure 3). Trembling aspen, however, showed a slight increase in reflectance at the 570 nm and the 800 nm regions (Table I). As expected, these minor changes were not detected from airborne MEIS II data.

CONCLUSIONS

Spectral reflectance measurements from these sites in the Canadian Shield revealed only small changes between mineralized and background sites. This finding suggests that the vegetation is not heavily stressed over these sites, despite relatively high geochemical values in the soil. Airborne detection of changes from geochemical stress is, therefore, difficult. In the case of the White Lake mine, where changes were significant, detection by high-resolution airborne sensors was possible.

REFERENCES

- Bailes, A.H., 1971. Preliminary Compilation of the Geology of the Snow Lake-Flin Flon, Sherridon Area. Manitoba Mines Branch.
- Carter, M.W., 1976. Kelvin Township, District of Sudbury, Ontario. Division of Mines, Preliminary Map #1105, Geological Series. Toronto, Ontario.
- Collins, W., S.H. Chang, G. Raines, F. Canney and R. Ashley, 1983. Airborne Biogeochemical Mapping of Hidden Mineral Deposits. *Economic Geology*, Vol.78, pp.737-749.
- Gladwell, D.R., R.E. Lett and P. Lawrence, 1983. Application of Reflectance Spectrometry to Mineral Exploration Using Portable Radiometers. *Economic Geology*, Vol. 78, pp.699-710.
- Hoffer, R., 1978. Biological and Physical Considerations in Applying Computer-Aided Analysis Techniques to Remote Sensor Data in Remote Sensing: The Quantitative Approach, edited by Swain and Davis, McGraw Hill, New York, pp.227-289.
- Horler, D.N.H., J. Barber, A.R. Barringer, 1980. Effects of Heavy Metals on the Absorbance and Reflectance Spectra of Plants. *International Journal of Remote Sensing*, Vol.1, pp.121-136.
- Horler, D.N.H., M. Dockray and J. Barber, 1983. The Red Edge of Plant Reflectance. *International Journal of Remote Sensing*, Vol.4, #2, pp.273-288.
- Labovitz, J.I., E.J. Masuoka, R. Bell, A.W. Siegrist, R.F. Nelson, 1983. The Application of Remote Sensing to Geobotanical Exploration of Metal Sulphides. *Economic Geology*, Vol.78, pp.750-760.
- Rock, B.N., T. Hoshizak, and J.R. Miller, 1988. Comparison of In-Situ and Airborne Spectral Measurements of the Blue Shift Associated with Forest Decline. *Remote Sensing of the Environment*, Vol.24, #1, pp.109-129.
- Schwaller, M.R., and S.J. Tkack, 1980. Premature Leaf Senescence as an Indicator in Geobotanical Prospecting with Remote Sensing Techniques. *Proceedings of the 14th International Symposium Remote Sensing of the Environment*, San Jose, Costa Rica, pp.347-358.
- Singhroy, V., R. Stanton-Gray, J. Springer, 1986. Spectral Geobotanical Investigations of Mineralized Till Sites. *Proceedings of the 5th Thematic Conference, Remote Sensing for Exploration Geology*, ERIM, Ann Arbor, Michigan, pp.523-543.
- Singhroy, V.H., 1987. An Overview of Geobotanical Remote Sensing in Canada. *Proceedings of the 11th Canadian Symposium on Remote Sensing*, Waterloo, Ontario, pp.259-275.
- Till, S.M., R.A. Neville, W.D. McColl, R.P. Gauthier, J.R. Miller, E.W. Hare, 1984. Application of MEIS II Airborne Pushbroom Scanner to Geobotanical Remote Sensing. *Proceedings of the 3rd Thematic Conference on Remote Sensing for Exploration Geology*, ERIM, Ann Arbor, Michigan, pp.885-888.

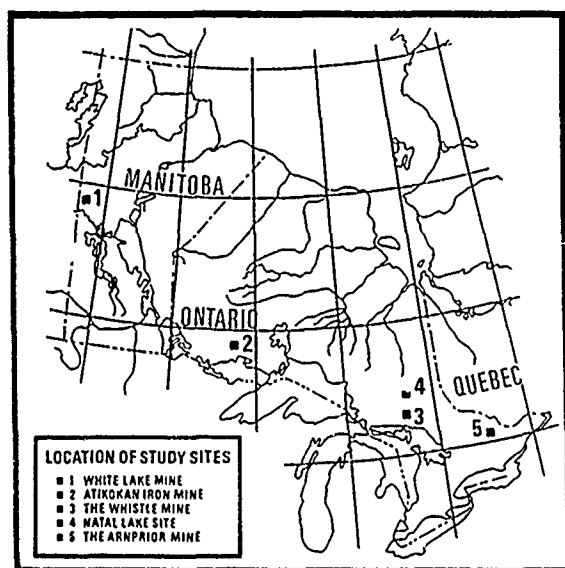
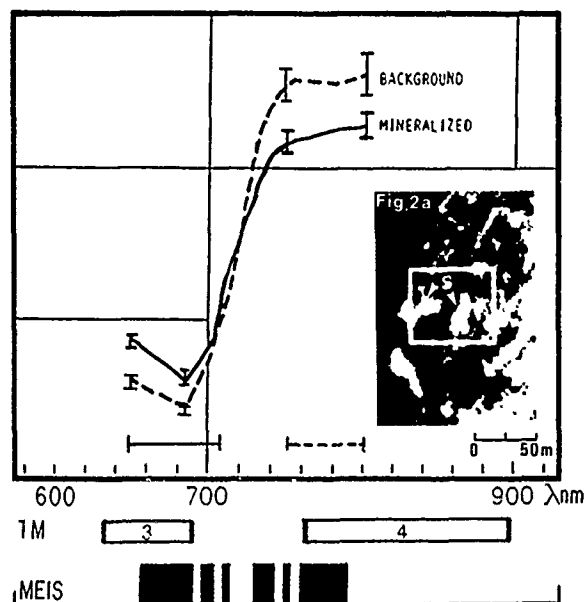
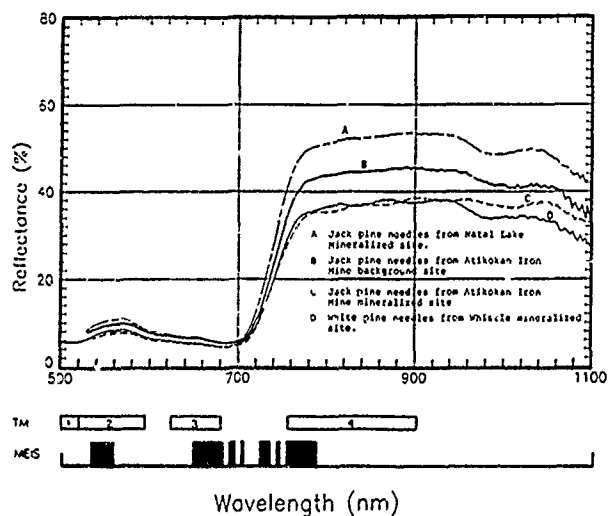


Figure 1



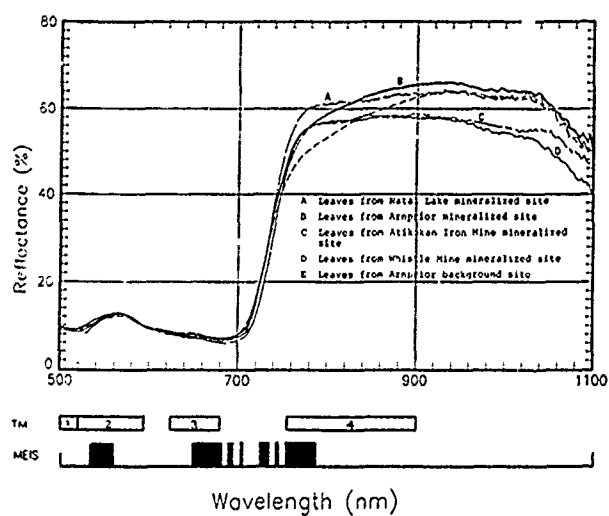
AVERAGE SPECTRAL REFLECTANCE OF JACK PINE NEEDLES AT THE WHITE LAKE MINE SITE. 2A SHOWS ENHANCED MEIS II IMAGERY DEPICTING GEOCHEMICALLY STRESSED VEGETATION AT S.

Figure 2



AVERAGE SPECTRAL REFLECTANCE CURVES OF JACK PINE AND WHITE PINE NEEDLES TAKEN FROM MINERALIZED AND BACKGROUND SITES.

Figure 3



AVERAGE SPECTRAL REFLECTANCE CURVES OF WHITE BIRCH LEAVES TAKEN FROM SEVERAL MINERALIZED & BACKGROUND SITES IN THE CANADIAN SHIELD.

Figure 4

SUMMARY DATA

| Site/Date | # of Samples | Ref. Local Max. | λ Local Max. | Ref. Local Min | λ Local Min | λ (nm) & Pt. of Inflection | Ref. (%) & Shoulder | Shape Parameter (nm) |
|----------------------------|--------------|-----------------|----------------------|----------------|---------------------|------------------------------------|---------------------|----------------------|
| 1. White Lake * -summer 77 | | | | | | | | |
| -Jack Pine | - M 10 | - | - | 13.0 | - | - | 45.0 | - |
| | - B 6 | - | - | 8.0 | - | - | 52.0 | - |
| -White Spruce | - M 11 | - | - | 13.8 | - | - | 44.0 | - |
| | - B 5 | - | - | 10.0 | - | - | 56.2 | - |
| -Trembling Aspen | - M 10 | - | - | 12.8 | - | - | 63.2 | - |
| | - B 5 | - | - | 10.0 | - | - | 67.1 | - |
| -White Birch | - M 10 | - | - | 15.9 | - | - | 60.0 | - |
| | - B 2 | - | - | 12.5 | - | - | 64.3 | - |
| 2. Atikokan 23/8/87 | | | | | | | | |
| -White Birch | - M 10 | 12.8 | 567.7 | 5.9 | 681.6 | 736.9 | 56.8 | 55.3 |
| | - B 14 | 13.6 | 568.0 | 6.4 | 682.7 | 735.8 | 55.1 | 53.0 |
| -Trembling Aspen | - M 10 | 15.7 | 567.1 | 7.4 | 681.4 | 737.8 | 60.7 | 56.5 |
| | - B 8 | 14.3 | 567.3 | 7.9 | 685.2 | 743.3 | 59.5 | 58.0 |
| -Jack Pine | - M 12 | 7.9 | 566.3 | 4.3 | 683.6 | 738.1 | 35.3 | 54.5 |
| | - B 7 | 10.2 | 568.7 | 5.3 | 685.3 | 739.4 | 43.5 | 54.1 |
| -Balsam Fir | - M 8 | 9.9 | 566.9 | 5.1 | 681.9 | 739.9 | 39.5 | 58.0 |
| | - B 9 | 9.7 | 564.9 | 5.3 | 686.2 | 745.7 | 43.4 | 59.5 |
| 3. Arnprior 30/7/87 | | | | | | | | |
| -Trembling Aspen | - M 20 | 16.9 | 563.9 | 8.4 | 678.6 | 732.2 | 67.1 | 53.5 |
| | - B 15 | 17.6 | 561.2 | 7.1 | 675.6 | 738.3 | 66.8 | 62.7 |
| -White Birch | - M 32 | 12.5 | 566.5 | 7.0 | 676.5 | 729.8 | 53.0 | 53.2 |
| | - B 11 | 12.9 | 564.1 | 6.9 | 674.6 | 732.9 | 58.6 | 58.3 |
| -Red Oak | - M 12 | 12.3 | 560.6 | 6.5 | 675.5 | 736.9 | 54.4 | 61.4 |
| | - B 11 | 14.8 | 560.6 | 6.9 | 672.2 | 736.7 | 71.8 | 64.5 |
| -White Oak | - M 11 | 11.3 | 561.9 | 5.7 | 675.2 | 738.1 | 52.6 | 62.9 |
| | - B 7 | 14.5 | 566.6 | 6.3 | 678.9 | 741.1 | 73.2 | 62.1 |
| 4. Whistle Mine 8/8/88 | | | | | | | | |
| -White Pine | - M 19 | 8.5 | 564.6 | 4.6 | 684.9 | 742.9 | 36.4 | 58.0 |
| | - B 18 | 9.1 | 562.7 | 4.6 | 680.9 | 744.2 | 41.9 | 63.2 |
| -White Birch | - M 18 | 12.9 | 564.9 | 6.7 | 679.9 | 740.2 | 56.9 | 60.3 |
| | - B 15 | 10.5 | 564.8 | 5.6 | 678.4 | 738.9 | 49.8 | 60.5 |
| 5. Natal 19/8/86 | | | | | | | | |
| -Trembling Aspen | - M 24 | 13.0 | 564.3 | 6.8 | 680.4 | 741.2 | 66.6 | 60.8 |
| | - B 25 | 12.6 | 563.2 | 6.4 | 677.9 | 738.1 | 62.6 | 60.2 |
| -Jack Pine | - M 11 | 11.1 | 566.3 | 5.4 | 681.9 | 739.7 | 50.9 | 57.8 |
| | - B 18 | 10.0 | 564.6 | 5.0 | 680.9 | 740.4 | 47.9 | 59.5 |
| -White Birch | - M 17 | 14.1 | 563.3 | 6.8 | 678.8 | 734.9 | 63.9 | 56.2 |
| | - B 25 | 14.2 | 564.4 | 7.2 | 675.9 | 736.2 | 68.3 | 60.3 |
| -Black Spruce | - M 6 | 10.2 | 560.6 | 6.5 | 681.9 | 739.6 | 40.7 | 57.7 |
| | - B 11 | 11.4 | 563.6 | 6.6 | 682.4 | 739.4 | 49.1 | 57.0 |

LEGEND

M - Mineralized Site

B - Background Site

1. # of samples - the number of samples in averages

2. Ref. Local Max - Reflectance (%) at local maximum near 570 nm

3. λ Local Max - Wavelength (nm) at local maximum near 570 nm

4. Ref. Local Min - Reflectance (%) at local minimum near 680 nm

5. λ Local Min - Wavelength (nm) at local minimum near 680 nm6. λ Point of Inflection - Wavelength (nm) at the Point of Inflection on the Red edge

7. Ref. Shoulder - Reflectance (%) at 800 nm

8. Shape Parameter - Wavelength at the Point of Inflection minus the wavelength at the local Minimum

* For the White Lake Site 1977 a Barringer Radiometer with 10 filters was used (650, 680, 685, 703, 713, 733, 743, 753, 782 and 800 nm), therefore the accuracy of locating the key shape parameters is not as accurate as using with the other sites where a spectroradiometer was used that records reflectance every 2.8 nm.

Table I

SUMMARY DATA

| Site/Date | # of Samples | Ref. Local Max. | λ Local Max. | Ref. Local Min | λ Local Min | λ (nm) * Pt. of Inflection | Ref. (%) * Shoulder | Shape Parameter (nm) |
|-----------------------------------|--------------|-----------------|----------------------|----------------|---------------------|------------------------------------|---------------------|----------------------|
| 1. White Lake * -summer 77 | | | | | | | | |
| -Jack Pine - M | 10 | - | - | 13.0 | - | - | 45.0 | - |
| -B - B | 6 | - | - | 8.0 | - | - | 52.0 | - |
| -White Spruce - M | 11 | - | - | 13.8 | - | - | 44.0 | - |
| -B - B | 5 | - | - | 10.0 | - | - | 56.2 | - |
| -Trembling Aspen - M | 10 | - | - | 12.8 | - | - | 63.2 | - |
| -B - B | 5 | - | - | 10.0 | - | - | 67.1 | - |
| -White Birch - M | 10 | - | - | 15.9 | - | - | 60.0 | - |
| -B - B | 3 | - | - | 12.5 | - | - | 64.3 | - |
| 2. Atikokan 23/8/87 | | | | | | | | |
| -White Birch - M | 10 | 12.8 | 567.7 | 5.9 | 681.6 | 736.9 | 56.8 | 55.3 |
| -B - B | 14 | 13.6 | 568.0 | 6.4 | 682.7 | 735.8 | 53.1 | 53.0 |
| -Trembling Aspen - M | 10 | 15.7 | 567.1 | 7.4 | 681.4 | 737.8 | 60.7 | 56.5 |
| -B - B | 8 | 14.3 | 567.3 | 7.9 | 685.2 | 743.3 | 59.5 | 58.0 |
| -Jack Pine - M | 12 | 7.9 | 566.3 | 4.3 | 683.6 | 738.1 | 35.3 | 54.5 |
| -B - B | 7 | 10.2 | 568.7 | 5.3 | 685.3 | 739.4 | 43.5 | 54.1 |
| -Balsam Fir - M | 8 | 9.9 | 566.9 | 5.1 | 681.9 | 739.9 | 39.5 | 58.0 |
| -B - B | 9 | 9.7 | 564.9 | 5.3 | 686.2 | 745.7 | 43.4 | 59.5 |
| 3. Annapolis 30/7/87 | | | | | | | | |
| -Trembling Aspen - M | 20 | 16.9 | 563.9 | 8.4 | 678.6 | 732.2 | 67.1 | 53.5 |
| -B - B | 15 | 13.6 | 561.2 | 7.1 | 675.6 | 738.3 | 66.8 | 62.7 |
| -M - M | 32 | 12.5 | 566.5 | 7.0 | 676.5 | 729.8 | 53.0 | 53.2 |
| -White Birch - B | 11 | 12.9 | 564.1 | 6.9 | 674.6 | 732.9 | 58.6 | 58.3 |
| -M - M | 12 | 12.3 | 560.6 | 6.5 | 675.5 | 736.9 | 54.4 | 61.4 |
| -Red Oak - B | 11 | 14.8 | 560.6 | 6.9 | 672.2 | 736.7 | 71.8 | 64.5 |
| -M - M | 11 | 11.3 | 561.9 | 5.7 | 675.2 | 738.1 | 52.6 | 62.9 |
| -B - B | 7 | 14.5 | 566.6 | 6.3 | 678.9 | 741.1 | 73.2 | 62.1 |
| 4. Whistle Mine 8/8/88 | | | | | | | | |
| -White Pine - M | 19 | 8.5 | 564.6 | 4.6 | 684.8 | 742.9 | 36.4 | 58.0 |
| -B - B | 18 | 9.1 | 562.7 | 4.6 | 680.9 | 744.2 | 41.9 | 63.2 |
| -M - M | 18 | 12.9 | 564.9 | 6.7 | 679.9 | 740.2 | 56.9 | 60.3 |
| -White Birch - B | 15 | 10.5 | 564.8 | 5.6 | 678.4 | 738.9 | 49.0 | 60.5 |
| 5. Natal 19/8/86 | | | | | | | | |
| -Trembling Aspen - M | 24 | 13.0 | 564.3 | 6.8 | 680.4 | 741.2 | 66.6 | 60.8 |
| -B - B | 25 | 12.6 | 563.2 | 6.4 | 677.9 | 738.1 | 62.6 | 60.2 |
| -M - M | 11 | 11.1 | 566.3 | 5.4 | 681.9 | 739.7 | 50.9 | 57.8 |
| -B - B | 18 | 10.0 | 564.6 | 5.0 | 680.9 | 740.4 | 47.9 | 59.5 |
| -White Birch - M | 17 | 14.1 | 563.3 | 6.8 | 678.8 | 734.9 | 63.9 | 56.2 |
| -B - B | 25 | 14.2 | 564.4 | 7.2 | 675.9 | 736.2 | 68.3 | 60.3 |
| -M - M | 6 | 10.2 | 560.6 | 6.5 | 681.9 | 739.6 | 40.7 | 57.7 |
| -Black Spruce - B | 11 | 11.4 | 563.6 | 6.6 | 682.4 | 739.4 | 49.1 | 57.0 |

LEGEND

M - Mineralized Site

B - Background Site

1. # of samples - the number of samples in averages

2. Ref @ local Max - Reflectance (%) at local maximum near 570 nm

3. λ @ local Max - Wavelength (nm) at local maximum near 570 nm

4. Ref @ local Min - Reflectance (%) at local minimum near 680 nm

5. λ @ local Min - Wavelength (nm) at local minimum near 680 nm6. λ @ Point of Inflection - Wavelength (nm) at the Point of Inflection on the Red edge

7. Ref of Shoulder - Reflectance (%) at 800 nm

8. Shape Parameter - Wavelength at the Point of Inflection minus the wavelength at the local Minimum

* For the White Lake Site 1977 a Barringer Radiometer with 10 filters was used (650, 680, 685, 703, 713, 733, 743, 753, 782 and 800 nm), therefore the accuracy of locating the key shape parameters is not as accurate as using with the other sites where a spectral radiometer was used that records reflectance every 2.8 nm.

Table I

BIDIRECTIONAL REFLECTANCES OF THREE SOIL SURFACES AND THEIR CHARACTERIZATION THROUGH MODEL INVERSION

D.W. Deering

T.F. Eck

J. Otterman

NASA/Goddard Space Flight Center
Earth Resources Branch
Greenbelt, MD 20771, USA
(301) 286-9186

ST Systems Corporation
Lanham, MD, USA
(301) 286-6559

Tel Aviv University
Ramat Aviv, Israel
(currently at GSFC,
Greenbelt, MD, USA)
(301) 286-7208

ABSTRACT

Spectral bidirectional reflectances were measured over three natural soil sites using a specially designed radiometer called the PARABOLA. Two of the sites were bare soils and the third had a sparse cover of desert scrub. The reflectances were strongly non-Lambertian for all three surfaces but with markedly different patterns. The measured data were fitted with a quasi-physical reflectance model in which the surface backscattering and forwardscattering are separately formulated. A soil reflectance characterization was obtained by assessing the contributions of the forward, backward and Lambertian components. This three-parameter characterization produced a satisfactory fit to the measured reflectances, and appears promising to provide a basis for soils categorization.

KEYWORDS: bidirectional reflectance, modeling, anisotropy, soil

INTRODUCTION

A unique two-axis scanning-head field radiometer, called the PARABOLA (Deering and Leone, 1986), has been used to measure the directional incoming and outgoing spectral radiances for a variety of earth surface types. Our study involves characterizing the bidirectional reflectances for three relatively simple surface types: two bare soils and a third surface with soil similar to one of the bare soils but with a sparse plant canopy. We then express the bidirectional reflectance patterns in terms of three soil reflectance components from inversion of a quasi-physical model. The goal herein is 1) to evaluate and characterize the nature of the bidirectional reflectance variations and 2) develop an approach to simple radiometric surface categorization for soils, both bare and with sparse vegetation.

SITE CHARACTERISTICS AND FIELD INSTRUMENTATION

The two bare soil surfaces studied are a naturally occurring dune sand and an alkali

flat in the White Sands desert (<250 mm annual precipitation) of New Mexico. A dune sand flat, whose surface is composed of sand-sized, nearly pure gypsum (hydrous calcium sulfate) crystals and is characterized by uniform ripples formed by wind action, was selected for the measurements (Fig. 1a and 1b). These ripples are, in effect, transverse "micro-dunes", which averaged approximately 10 cm from crest to crest and approximately 1-2 cm from crest to trough.

The alkali flat examined occurs within a vast non-vegetated expanse (32 km long) of the Tularosa Basin along the predominantly windward side of the White Sands dune formation. The alkali (or gypsum) flat's surface microtopography is coarse-textured and is composed of a stabilized crust with patches of a darker coloration.

The third surface type is a low-scrub community that lies within a semidesert grassland area in Ward County in west Texas (Fig. 1c). With a long history of overgrazing, the vegetation cover consisted primarily of scattered clumps of a low growing Rough Coldenia subshrub species (*Coldenia hispidissima*) mixed with several other scrub species and grasses of a similar stature. The 10-15 cm tall, grayish-woody shrub had some green leaves and the vegetative clumps occupied less than 15 percent of the surface area.

The Portable Apparatus for Rapid Acquisition of Bidirectional Observations of the Land and Atmosphere, or PARABOLA instrument has been described in detail by Deering and Leone (1986). The PARABOLA is a battery-powered, two-axis scanning head three-channel (0.650-0.670, 0.810-0.840, and 1.620-1.690 μm , respectively), motor-driven radiometer that enables the acquisition of radiance data for almost the complete (4π) sky- and ground-looking hemispheres in 15° instantaneous field-of-view (IFOV) sectors in only 11 s. The instrument was operated at 4.5 m above the ground surface.

THE SURFACE BIDIRECTIONAL REFLECTANCE MODEL

The bidirectional reflectance model specifies separately the contribution of the soil and from the plants. The reflectance of bare soil, R_s , is given as:

$$R_s = (1-f) \frac{r(\sin z - z \cos z) + t[(z-\pi)\cos(z-\pi) - \sin(z-\pi)]}{4(\cot \theta_o + \cot \theta_v)} + fr_o \quad (1)$$

where z is the azimuth with respect to the solar principal plane ($z = 0$ corresponds to forwardscattering in the principal plane), θ_o is the solar zenith angle and θ_v is the view zenith angle. The model is quasi-physical, as it describes a surface consisting of a Lambertian fraction, f , with reflectance r_o and a

predominant in backscattering, while the transmittance is predominant in forwardscattering. R_s represents the reflection from the soil unobscured by plants. The soil characterization is by three components: fr_o , $(1-f)r$ and $(1-f)t$.

The contribution of the plants is described by the same model of vertical cylinders or randomly oriented facets

$$R_p = \frac{r_p(\sin z - z \cos z)}{4(\cot \theta_o + \cot \theta_v)} \quad (2)$$

where r_p is the plant material reflectance. This model (Otterman and Weiss, 1984) was based on observations of desert-scrub vegetation in the Sinai (Otterman, 1981). The above expression for R_p represents a dense field of vertical cylinders. For a finite density, given by projection s of cylinders per unit area on a vertical plane, the contribution is multiplied by $1 - \exp[-s(\tan \theta_o + \tan \theta_v)]$. The effects of soil obscuration and soil shadowing by the plants are expressed by $\exp[-s(\tan \theta_o + \tan \theta_v)]$. The combined reflectance R_c of plants/soil is thus:

$$R_c = R_s(\exp[-s(\tan \theta_o + \tan \theta_v)] + R_p(1 - \exp[-s(\tan \theta_o + \tan \theta_v)])) \quad (3)$$

RESULTS

Reflectance Measurements

The measured bidirectional reflectances are presented in Figure 2. The patterns are quite different for the three soil surfaces, with the backscattering predominant for the alkali flat and for the desert scrub, but forwardscattering is dominant for the dune sand flat. There is considerable solar zenith angle dependence for the dune flat in the forwardscatter direction. For the desert scrub the solar zenith angle dependence is especially pronounced in the backscatter direction at large viewing zenith angles. Because the measurements at the White Sands were made in November, the range of solar zenith angles available was quite limited; theoretically, from 90° up to a solar noon position of about 52° . Measurements at the Texas site on September 14, however, were possible up to 29° .

The near-infrared ($0.826 \mu\text{m}$) reflectances (not presented here) followed, for all three surface types, very similar patterns of anisotropy to those in the red band, with typically 5-10% higher (absolute) values than the red reflectances. The shortwave infrared ($1.658 \mu\text{m}$) reflectances, on the other hand, were generally 10-20% lower (absolute) and exhibited somewhat stronger anisotropy than the red band.

Model Results

The quasi-physical bidirectional reflectance model specifies separately the contribu-

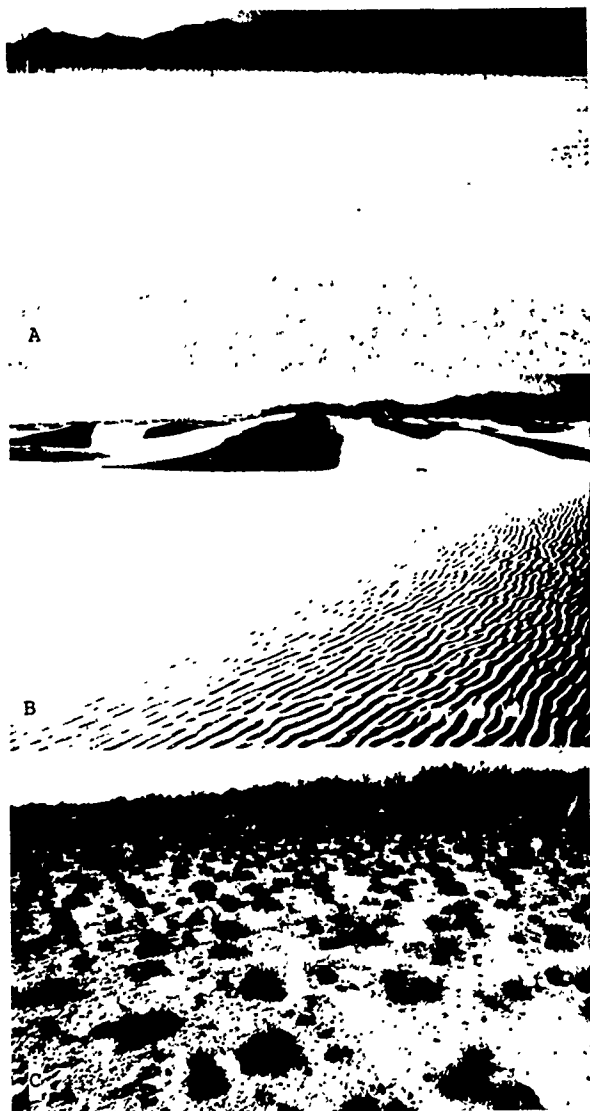


Figure 1. Photographs of the three alkali soil sites examined: a) alkali flat, b) dune sand flat and c) desert scrub site.

fraction $1-f$ of randomly-located vertical facets with facet-reflectance r and facet transmittance t . The reflection from such facets is equivalent to that from thin vertical cylinders, as described by Otterman and Weiss (1984). The facet reflectance is

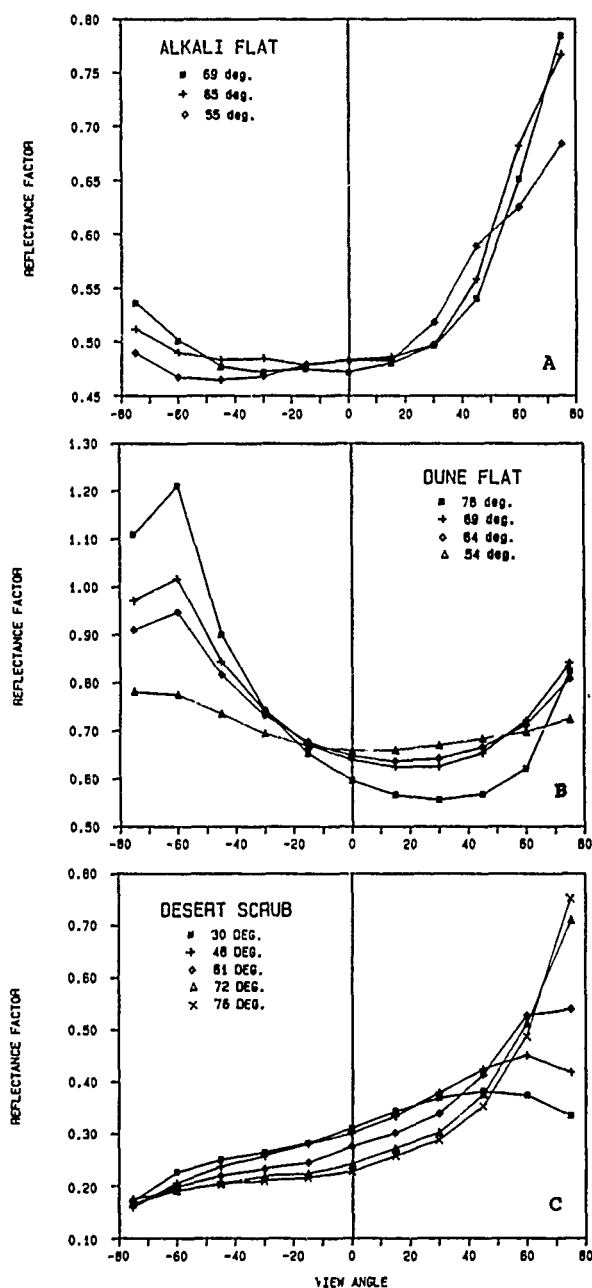


Figure 2. Red spectral band ($0.662\mu\text{m}$) reflectances in the solar principal plane at several solar zeniths for the alkali flat (a), dune sand flat (b), and the desert scrub (c) surfaces. Negative view angles are forwardscatter.

tion from the soil and from the plants. The enhanced anisotropy (both forwardscatter and backscatter) at large solar zenith angles that we observed is appropriately described by our model by $\cot \theta_s$ that appears in the denominator. The parameter values presented in Table 1 quantify the essential differences between the three surface types in terms of the Lambertian, backscatter, and

forwardscatter model components. The Lambertian component values for the alkali flat, dune sand and desert scrub surfaces remain essentially unchanged with solar zenith angles

Table I. Model inversion for each of the three soil surfaces made with approximately 55 observations from all view zenith angles $<70^\circ$ on one side of the solar principal plane, including the solar principal plane.

| θ° | f_r | $(1-f)r$ | $(1-f)t$ |
|----------------|-------|----------|----------|
| DUNE SAND FLAT | | | |
| 69° | 0.62 | 0.14 | 0.32 |
| 54° | 0.64 | 0.09 | 0.19 |
| ALKALI FLAT | | | |
| 69° | 0.41 | 0.26 | 0.09 |
| 55° | 0.44 | 0.30 | 0.05 |
| DESERT SCRUB | | | |
| 72° | 0.29 | 0.62 | 0.06 |
| 46° | 0.31 | 0.59 | 0.00 |

and typify the general magnitude of the reflectances at nadir. Thus, at the higher sun elevation examined the dune sand had the highest value at 0.64, the alkali flat was somewhat less at 0.44 and the desert scrub was the lowest at 0.31. The consistency of these three parameter values over a wide range of solar zenith angles is illustrated for the desert scrub type in Table 2 for both the red and near-infrared spectral bands.

The ratio of the backscattering component to the forwardscattering component is one indicator of the anisotropy. For the desert scrub the backscatter was larger than the forwardscatter by at least a factor of ten

Table II. Model inversion for the desert scrub surface over a wide range of solar zenith angles for the red and near-infrared spectral bands.

| θ° | f_r | $(1-f)r$ | $(1-f)t$ | g |
|-------------------------------|-------|----------|----------|------|
| 0.662 μm (RED) | | | | |
| 29.9° | .31 | .56 | .00 | .155 |
| 46.0° | .31 | .59 | .00 | .135 |
| 61.0° | .29 | .62 | .04 | .125 |
| 72.0° | .29 | .62 | .06 | .120 |
| 0.826 μm (NEAR-IR) | | | | |
| 29.9° | .38 | .55 | .04 | .125 |
| 46.0° | .39 | .62 | .02 | .125 |
| 61.0° | .37 | .60 | .02 | .100 |
| 72.0° | .33 | .54 | .04 | .070 |

over the wide range of solar zeniths examined. The alkali flat ratio was lower and varied with solar zenith angle; with ratio factors of 2.8 and 6.6 for 69° and 55°, respectively. The same ratio for the dune sand, with its strong specular reflectance, resulted in fractional ratio values of 0.43 and 0.47 (69° and 54°, respectively); and thus was much less dependent on solar zenith angle. Red and near-infrared spectral bands yielded similar

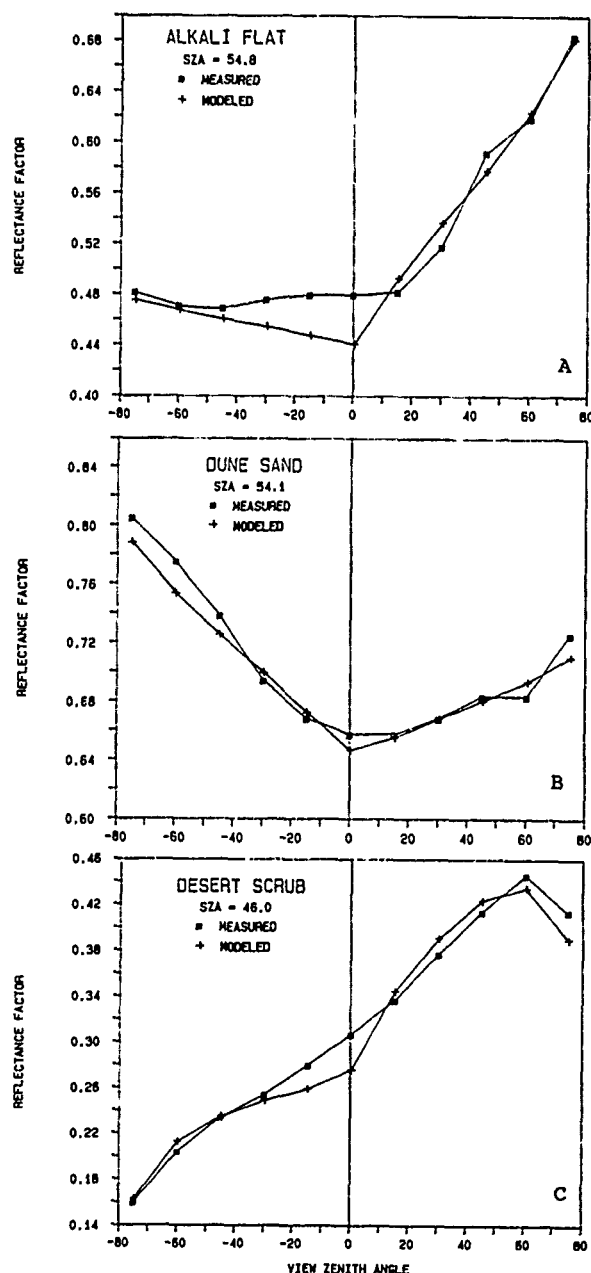


Figure 3. Measured and modeled bidirectional spectral reflectances at $0.662 \mu\text{m}$ for the a) alkali flat, b) dune sand and c) desert scrub surfaces for the nominal 50° solar zenith angle.

ratio values, although the Lambertian component was more variable with solar zenith angle for the near-infrared than for the red spectral band.

The correspondence of model-derived values of red spectral band bidirectional reflectance factors to measured values is presented for the solar principal plane in Figure 3. The curves compare very well for the dune sand and desert scrub surfaces, but show an appreciable deviation from the alkali

flat at the nadir and near-nadir forwardscatter viewing directions. With the considerable variety of curve shapes, the comparability is quite encouraging. Our model appears to provide a simple and very useful descriptive tool for directional reflectance analysis.

DISCUSSION AND CONCLUSIONS

The measured reflectances represented in our field program showed very strong non-Lambertian characteristics. The patterns of directional reflectance were sharply different for the two bare soils. The presence of sparsely placed plants on a relatively smooth soil also exhibits strong anisotropy, with plant shadowing affecting the anisotropy in a special way -- reducing the reflectance at large view zenith angles. For the desert scrub alkali soil site the ratio of the backscattering component to the forwardscattering component was larger by at least a factor of ten. The alkali flat ratio was somewhat lower and depended on solar zenith angle. The same ratio for the dune sand, with its strong specular reflectance, was close to 0.5 and was only slightly dependent on solar zenith angle. Red and near-infrared spectral bands yielded similar ratio values, although the Lambertian component was more variable with solar zenith angle for the near-infrared than for the red spectral band.

With only three surface parameters, the model inversion provides a satisfactory description of the surface spectral reflectance characteristics. The only significant departures from the modeled and the measured reflectances occurred in the vicinity of the nadir direction. Therefore, specifying the three anisotropy components (back scattering, forwardscattering, and Lambertian), appears to be a suitable approach to the characterization of soil type from the point of view of bidirectional reflectance patterns. Some refinement of the model might be appropriate, and additional bare soil and sparsely-covered soil surface measurements will be made to further develop and validate this approach.

One important value of the model is that relatively minor changes in this quantitative characterization occur when the solar zenith angle changes. It is anticipated that based on this characterization useful but simple radiometric surface categorization for soils, both bare and with sparse vegetation, may be developed.

REFERENCES

- Deering, D.W. and P. Leone, "A sphere-scanning radiometer for rapid directional measurements of sky and ground radiance", *Remote. Sens. Environ.* 19: 1-24, 1986.
- Otterman, J. "Satellite and field studies of man's impact on the surface in arid regions", *Tellus*. 33: 68-77, 1981.
- Otterman, J. and G.H. Weiss, "Reflections from a field of randomly located vertical protrusions", *Appl. Opt.* 23(12): 1931-1936, 1984.

THE VARIOUS CHLOROPHYLL FLUORESCENCE SIGNATURES AS A BASIS FOR PHYSIOLOGICAL GROUND TRUTH CONTROL IN REMOTE SENSING OF FOREST DECLINE

U. Rinderle and H.K. Lichtenthaler

Botanical Institute II (Plant Physiology and Plant Biochemistry)
University of Karlsruhe, Kaiserstr. 12, D-7500 Karlsruhe (FRGermany)
Tel.: 0049/721/608-3833 Telefax: 721/608 4290

ABSTRACT

Different parameters of the in vivo chlorophyll fluorescence were measured from August 1987 to December 1988 in the needles of two spruce sites in the Northern Black Forest (Althof, damage class 0/1, and Mauzenberg, damage class 3/4). The fluorescence parameters measured comprise the Kautsky induction kinetics with Rfd-values as vitality index (at 690 and 730 nm), the photochemical and non-photochemical quenching coefficients (qQ and qE), the ratio of maximum to ground fluorescence Fm/Fo, the height of the saturation fluorescence peak (g-h) as well as the ratio of the two fluorescence-emission maxima F690/F735. The appearance of stress and damage as well as regeneration of the needles, as visualized from the seasonal changes of these chlorophyll-fluorescence parameters, are described in detail. The differences in the various fluorescence signatures and in photosynthetic activity (CO₂-assimilation rates) between needles of healthy and damaged trees are discussed with respect to stress detection and ground-truth control in connection with remote sensing of terrestrial vegetation.

Key words: chlorophyll fluorescence, Rfd-values, ratio F690/F735, photosynthetic activity, forest decline.

1. INTRODUCTION

Red chlorophyll fluorescence emitted from green plant tissue can be taken as an instrument to obtain an insight into the processes of photosynthesis. The inverse relationship between in vivo chlorophyll fluorescence and photosynthetic activity can be used to study the potential photosynthetic capacity of plants as well as to detect damage and stress. Measurements of fluorescence signatures are non-destructive and can be applied for physiological ground-truth control (Buschmann et al., 1989; Lichtenthaler et al., 1986; Schmuck et al., 1987). There are several fluorescence parameters which contain different physiological information.

The light-induced in vivo chlorophyll fluorescence of a predarkened leaf shows a transient which is known as fluorescence induction kinetic (Kautsky effect). The fluorescence-decrease ratio (Rfd = f_d/f_s = fluorescence decrease to steady-state fluorescence) has been established as a vitality index of leaves (Lichtenthaler et al., 1986; Lichtenthaler and Rinderle, 1988a,b,c). The height of the Rfd-values (measured in the 690 and 735 nm regions with the two-wavelength fluorometer) reflect the potential photosynthetic activity of leaves. Though Rfd-values usually go parallel with the net CO₂-assimilation rates, they are a different parameter, and are independent of the stomata opening, signaling even at closed stomata whether the photosynthetic apparatus is photochemically active or not (Lichtenthaler, 1988).

The chlorophyll-fluorescence emission spectra measured at room temperature exhibit two maxima in the 690 and 735 nm regions. The measurement of the spectra allows determination of the ratio of the two fluorescence maxima (ratio F690/F735). The level of this ratio not only reflects the chlorophyll content of the needles, but also their photosynthetic activity (Rinderle and Lichtenthaler, 1988a,b). In a provisional airborne system equipped with a pulsed excimer laser ($\lambda = 450$ nm) as excitation light, differences in the chlorophyll-fluorescence ratio F690/F735 between trees of different physiological state were detected (Zimmermann and Günther, 1986). This ratio may be the basis for a LIDAR-system to sense the state of physiology of terrestrial vegetation (Lichtenthaler, 1989).

The PAM-fluorometer permits determination of the photochemical Q- and the non-photochemical E-quenching of chlorophyll fluorescence and the rate of the Q_A-reoxidation in the photosynthetic electron transport chain (g-h), as well as the ratio of the maximum to ground fluorescence Fm/Fo (Schreiber et al., 1986; Lichtenthaler and Rinderle, 1988a).

The seasonal variation in the various chlorophyll-fluorescence signatures, in the pigment content (chlorophylls and carotenoids) and in photosynthetic activity (CO₂-assimilation rate P_g) between needles of healthy and damaged spruce trees (PICEA ABIES) of the Northern Black Forest during the spring, summer and winter period was compared in order to determine which parameters are best suitable for physiological ground-truth control.

2. MATERIAL AND METHODS

The development of fluorescence signatures of mainly healthy (Althof, damage class 0/1, normal green needles, little needle loss) and of damaged spruce trees (Mauzenberg, damage class 3/4, 70 to 80 % needle loss, yellowish-green needles) were determined during 17 months using three different fluorescence methods:

1) The chlorophyll fluorescence emission spectra (at room temperature) induced by blue light (470 ± 30 nm) were recorded with a Shimadzu MPS 5000 spectrometer under steady-state conditions of the chlorophyll fluorescence (5 min after onset of illumination) and the ratio F690/F735 calculated (Rinderle and Lichtenthaler, 1988).

2) The red laser-induced chlorophyll fluorescence kinetics (determination of Rfd-values as vitality index of needles) measured simultaneously at 690 and 730 nm in a portable two-wavelength field fluorometer (Lichtenthaler and Rinderle, 1988a,c).

3) The differentiation between photochemical Q-quenching and non-photochemical E-quenching, the height of the saturation pulse-induced fluorescence spikes (g-h) and the ratio of maximum to ground fluorescence Fm/Fo were determined

using the PAM-fluorometer of Walz (Schreiber et al., 1986; Lichtenthaler and Rinderle, 1988a).

The photosynthetic prenyl pigments (chlorophylls and carotenoids) were extracted with 100% acetone and the pigments quantitatively determined using the new extinction coefficients (Lichtenthaler, 1987).

The CO_2 -assimilation (P_N), transpiration and stomatal conductivity (gH_2O) were determined using the $\text{CO}_2/\text{H}_2\text{O}$ -porometer system of Walz (Schulze et al., 1982).

2. RESULTS AND DISCUSSION

Pigment content:

The chlorophyll content of needles was followed from August 1987 to December 1988. The chlorophyll a+b content (per needle area unit) of the youngest and the older needle years of damaged trees (PICEA ABIES) was always lower than that of healthy spruces. With beginning of the vegetation period in May the chlorophyll content of all needle years increased at both spruce sites. In the case of the older needles of the damaged spruces, the chlorophyll content, however, decreased again during summer 1988 and reached its lowest values in the winter months (Fig. 1).

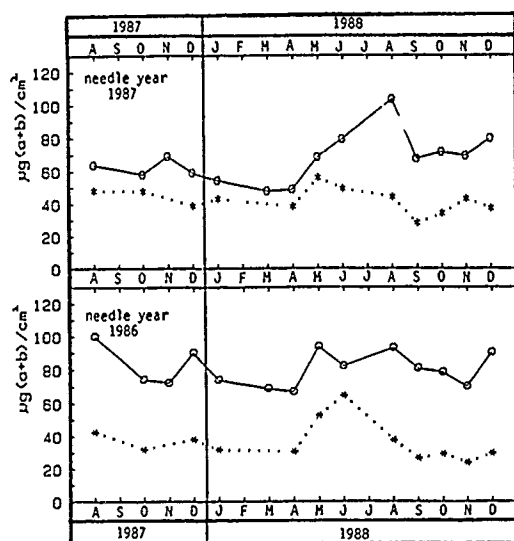


Fig. 1: Seasonal variation of the chlorophyll content (in μg (a+b) per cm^2 needle area) of two different needle years of healthy (o—o Althof; damage class 0/1) and damaged spruces (..... Mauzenberg; damage class 3/4) starting in August 1987. The large letters below and above the figure represent the months August 1987 to December 1988.

The needles of the damaged trees showed higher values for the ratio of chlorophyll a/b than the needles of the healthy trees, which indicates a lower amount of light harvesting chlorophyll a/b proteins in the needles of the damaged spruces. For the pigment ratio chlorophylls to carotenoids $(a+b)/(x+c)$ we found lower values throughout the year for the older needles of the damaged spruces than for those of healthy trees, except for the regreening period in spring. The lower values for the pigment ratio $(a+b)/(x+c)$ reflect longlasting stress events of the needles of the damaged trees.

Fluorescence ratio F690/F735:

The shape of the blue-light induced emission-spectra of green leaves or needles is characterized by two maxima in the 690 and 735 nm regions. In normal green needles a large proportion of the in vivo chlorophyll fluorescence in the 690

nm region is reabsorbed by the leaf chlorophylls, since the absorption bands of the latter overlap with the emitted fluorescence in this region. With decreasing chlorophyll content, the ratio of the two fluorescence maxima, the ratio F690/F735, increases because the probability of reabsorption is much lower in needles with a lower chlorophyll content. The light-green, not yet fully developed needles of the youngest needle year 1988, not shown here, exhibited higher values for the ratio F690/F735 (ca. 1.5-2), whereas green needles showed values of 0.9-1.0 (Fig. 2). The older yellowish-green needles of the damaged spruces also possess a lower chlorophyll content and consequently higher values for the ratio F690/F735 (1.2-1.6) than the fully green needles from the healthy trees. These differences between needles of healthy and damaged spruces were larger in autumn and winter than during the new accumulation of chlorophylls in May and June (Fig. 2). The ratio F690/F735 is therefore a very suitable stress indicator of the stress-induced lower chlorophyll content without the performance of pigment analysis. The values of F690/F735 are higher when the chlorophyll fluorescence is induced by blue light than by red light. Yet in both cases the values of this fluorescence ratio increase with decreasing chlorophyll content and photosynthetic function.

In leaves treated with the herbicide diuron (DCMU), which blocks the photosynthetic electron transport chain, we also observed an increase of the ratio F690/F735 though the chlorophyll content remained the same before and after the treatment. This indicates that not only the reduced chlorophyll content of the needles of the damaged trees increases the values of the ratio F690/F735, but also the loss of the photosynthetic function, as induced here by diuron (Lichtenthaler and Rinderle, 1988b).

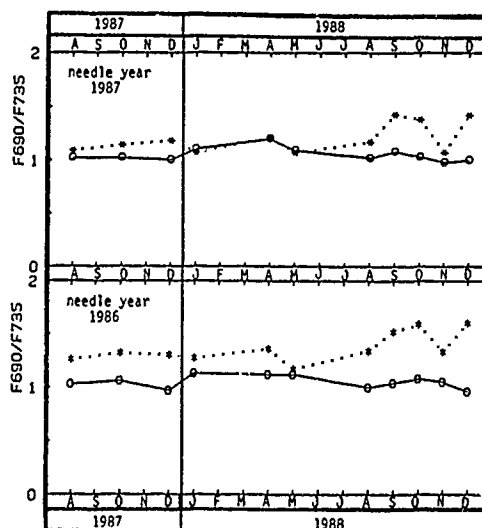


Fig. 2: Seasonal variation of the ratio of the fluorescence maxima, ratio F690/F735 of two different needle years of healthy (o—o Althof; damage class 0/1) and damaged spruces (..... Mauzenberg; damage class 3/4) starting in August 1987.

Rfd-values as vitality index:

From the slow component of the fluorescence induction kinetics of predarkened needles, one can determine the ratio of the fluorescence decrease (fd) to the steady-state fluorescence (fs). The values of this ratio ($R_{fd} = fd/fs$) are a valuable indicator of the potential photosynthetic activity of leaves and of a plant's vitality (Lichtenthaler and Rinderle, 1988a).

In autumn and winter the needles of the damaged trees showed lower R_{fd} -values than those of the healthy trees (Fig. 3).

During the formation of the new shoots in spring, however, the Rfd-values of the needles of damaged and healthy trees no longer exhibit differences. In December 1987, with a very strong frost period, the Rfd-values were considerably decreased; they showed a recovery in a relatively warm January 1988, however, and decreased again in a cold March. These characteristics were found in the needles at the Althof and Mauzenberg sites (Fig. 3). The decrease of the Rfd-values in December and March indicated a damage to the photosynthetic apparatus, the increase in January (warm period) demonstrated the fast regeneration of the photosynthetic function. The investigation thus demonstrates that the Rfd-values are a suitable parameter for screening the seasonal variation in the physiology of the photosynthetic apparatus as well as the differences between healthy and damaged trees.

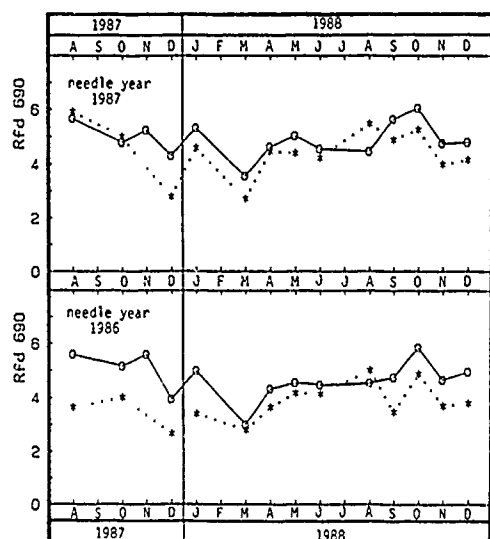


Fig. 3. Seasonal changes of the Rfd-values at 690 nm of two different needle years of healthy (o—o Althof; damage class 0/1) and damaged spruces (*... Mauzenberg; damage class 3/4) starting in August 1987.

From the Rfd-values at 690 and 730 nm measured with the two-wavelength field fluorometer one can determine the stress-adaptation index Ap (Strasser et al., 1987; Lichtenthaler and Rinderle, 1988a). Green, photosynthetically active needles of healthy spruces showed Ap -values of ca. 0.2-0.3. The Ap -values of the needles of the damaged trees were most of the time lower as those of the healthy ones. For Ap -values beneath 0.1 there is usually no possibility of regeneration of the needles. Such low Ap -values were only found in dying branches of damaged trees.

From the He/Ne-laser-induced fluorescence kinetics at 690 and 730 nm one can not only calculate the Rfd-values and the stress-adaptation index Ap , but from the steady-state fluorescence f_s in the 690 and 730 nm region also the ratio F_{690}/F_{735} at f_s . This ratio calculated from the induction kinetics has the same significance as the ratio F_{690}/F_{735} (calculated from the fluorescence emission spectra) described above.

CO₂-assimilation rate P_N :

The photosynthetic activity per-needle area unit (P_N , determined with a CO₂/H₂O-porometer system) was measured at light saturation and depended much on the stomata opening. At a stomatal conductivity (g_{H_2O}) with values of about 20-30 ($mmol\ H_2O \cdot m^{-2} \cdot s^{-1}$) good net CO₂-assimilation rates were observed and the stomata seemed to be reasonably far open.

Under water stress conditions and in wintertime (December 1987) the stomata were closed, resulting in very low CO₂-assimilation rates (Fig. 4). The relatively high Rfd-values at that time indicate that the internal photosynthetic apparatus was yet functional (assimilation of respiration CO₂). In autumn 1987 (August to October) the needles of the damaged trees showed lower values for the CO₂-assimilation rate than the needles of the healthy spruces. Before the formation of the new shoots in spring 1988 the healthy needles of the needle year 1986 exhibited high values for the CO₂-fixation, but during the summer and autumn of 1988 the net CO₂-rates decreased and were more or less the same level as those of the damaged needles (Fig. 4). In that time the Rfd-values of the healthy trees were slightly higher, indicating that the internal photosynthetic activity was functioning better than in the damaged trees.

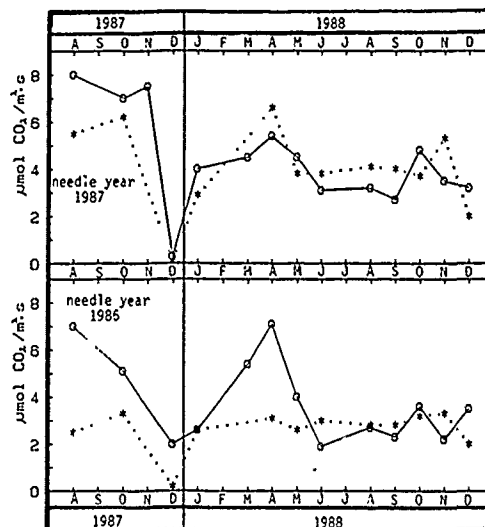


Fig. 4. Seasonal variation in the net CO₂-assimilation rate P_N (in $\mu mol\ CO_2 \cdot m^{-2} \cdot s^{-1}$) of two different needle years of healthy (o—o Althof; damage class 0/1) and damaged spruces (*... Mauzenberg; damage class 3/4) starting in August 1987.

In contrast to the P_N rates per needle area unit the net CO₂-assimilation rate per chlorophyll content ($mg\ CO_2 \cdot mg\ (a+b)^{-1} \cdot h^{-1}$) showed higher values for the needles of damaged trees, indicating that the small amounts of chlorophyll in the damaged needles were fully photosynthetically active.

Fluorescence measurements with the PAM-fluorometer:

Fluorescence spikes (g-h): With the PAM-fluorometer one can determine the chlorophyll fluorescence kinetics with short saturating light pulses given every 10 seconds. The pulse-induced fluorescence spikes (g-h) (Lichtenthaler and Rinderle, 1988a) indicate the Q_A -reoxidation capacity and were higher for normal green needles of the Althof site than those of the Mauzenberg site, not shown here. These differences are present in the one-year old and older needle years throughout the year, except for the regreening period in spring.

The quenching coefficients qQ and qE : From the original PAM-kinetics we calculated the quenching coefficients for the photochemical (qQ) and non-photochemical quenching (qE) (see Schreiber et al., 1986). The qQ -values of needles of healthy and damaged spruces were more or less the same (0.8-0.9) and did not vary during the course of the year.

The qE -values, however, which indicate the energy-dependent quenching of the absorbed light, showed in needles of healthy trees low values (0.3-0.4) in times of good photosynthetic activity, but were high in winter-time (0.5-0.6) (Fig. 5). In contrast, the older needles of the damaged spruces already

showed high values for qE in August 1987, through the winter and again in August 1988, which was probably due to a combined water, heat and light stress (Rinderle et al., 1988).

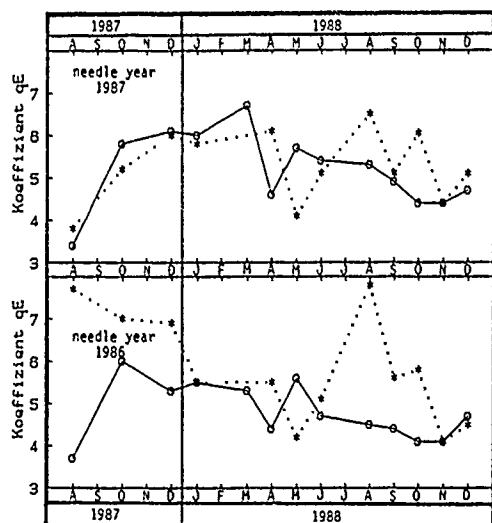


Fig. 5. Seasonal changes in the quenching coefficient qE of two needle years of healthy (o—o Althof; damage class 0/1) and damaged spruces (..... Mauzenberg; damage class 3/4) starting in August 1987. The qE -values were measured after a 10 min saturation pulse kinetic.

Ratio F_m/F_0 : The ratio of maximum fluorescence F_m to the ground fluorescence F_0 , as determined with the PAM-fluorometer, showed values around 5 in green spruce needles during summer-time; in winter values around 4 were detected. Needles of damaged trees showed somewhat lower values for this ratio, except for the regreening period in spring. Though the ratio contains some physiological information, it is not fully representative for the photosynthetic condition of the needles, since it is measured in the non-functional state 1 of photosynthesis.

4. CONCLUSION

The results of this comparative investigation of different needle years of mainly both healthy and damaged spruce trees indicate that the portable two-wavelength field-fluorometer and the PAM-fluorometer offer very good possibilities for ground-truth control measurements. The chlorophyll-fluorescence methods, especially the measurement of Rfd -values, the ratio F_{690}/F_{735} , the saturation pulse-induced fluorescence spikes (g-h) and the quench coefficient qE - in addition to measurements of chlorophyll content, pigment ratios and of net CO_2 -assimilation rate P_N - are very suitable criteria to judge physiology, stress appearance and partial regeneration of the photosynthetic apparatus in spring.

Acknowledgements: Part of this work was sponsored by the PEF, Karlsruhe which is gratefully acknowledged.

5. REFERENCES

Buschmann, C. Rinderle, U. and Lichtenthaler, H.K., "Detection of stress of coniferous forest trees with the VIRAF spectrometer", Internat. Geoscience and Remote Sensing Symposium, IGARSS Vancouver 1989, (these proceedings).

Lichtenthaler, H.K., "Chlorophylls and carotenoids, the pigments of photosynthetic biomembranes", Methods in Enzymology 148, pp. 350-382, 1987.

Lichtenthaler, H.K., "In vivo chlorophyll fluorescence as a tool for stress detection in plants", Applications of Chlorophyll Fluorescence, Lichtenthaler, H.K., ed., pp. 129-142, Kluwer Academic Publishers, Dordrecht 1988c.

Lichtenthaler, H.K., "Possibilities for remote sensing of terrestrial vegetation by a combination of reflectance and laser-induced chlorophyll fluorescence", Internat. Geoscience and Remote Sensing Symposium, IGARSS Vancouver 1989, (these proceedings).

Lichtenthaler, H.K. and Rinderle, U., "The role of chlorophyll fluorescence in the detection of stress conditions in plants", CRC Critical Reviews in Analytical Chemistry 19, Suppl. I, pp. S29-S85, 1988a.

Lichtenthaler, H.K. and Rinderle, U., "Chlorophyll fluorescence spectra of leaves as induced by blue light and red laser light", 4th Internat. Colloquium on Spectral Signatures of Objects in Remote Sensing, Aussois, pp. 251-254, ESA Publications Division, Noordwijk 1988b.

Lichtenthaler, H.K. and Rinderle, U., "Chlorophyll fluorescence as vitality indicator in forest decline research", Applications of Chlorophyll Fluorescence, Lichtenthaler, H.K., ed., pp. 143-149, Kluwer Academic Publishers, Dordrecht 1988c.

Lichtenthaler, H.K., Buschmann, C., Rinderle, U. and Schmuck, G., "Application of chlorophyll fluorescence in ecophysiology", Radiation Environm. Biophysics 25, pp. 297-308, 1986.

Rinderle, U. and Lichtenthaler, H.K., "The chlorophyll fluorescence ratio F_{690}/F_{735} as a possible stress indicator", Applications of Chlorophyll Fluorescence, Lichtenthaler, H.K., ed., pp. 189-196, Kluwer Academic Publishers, Dordrecht 1988.

Rinderle, U., Haitz, M., Lichtenthaler, H.K., Kähny, D.H., Shi, Z. and Wiesbeck, W., "Correlation of radar reflectivity and chlorophyll fluorescence of forest trees", Internat. Geoscience and Remote Sensing Symposium, IGARSS Edinburgh, Vol. 3, pp. 1343-1346, ESA Publications Division, Noordwijk 1988.

Schmuck, G., Lichtenthaler, H.K., Kritikos, G., Amann, V. and Rock, B., "Comparison of terrestrial and airborne reflection measurements of forest trees", Internat. Geoscience and Remote Sensing Symposium, IGARSS Michigan, Vol. 2, pp. 1201-1206, University of Michigan, Ann Arbor 1987.

Schreiber, U., Schliwa, U. and Bilger, W., "Continuous recording of photochemical and non-photochemical chlorophyll fluorescence quenching with a new type of modulation fluorometer", Photosynth. Res. 10, pp. 51-62, 1986.

Schulze, E.D., Hall, A.E., Lange, O.L. and Walz, H., "A portable steady state porometer for measuring the carbon dioxide and water vapour exchange of leaves under natural conditions", Oecologia 53, pp. 141-154, 1982.

Strasser, R., Schwarz, B. and Bucher, J., "Simultane Messung der Chlorophyllfluoreszenz Kinetik bei verschiedenen Wellenlängen als rasches Verfahren zur Frühdiagnose von Immissionsbelastungen an Waldbäumen", Europ. J. Forest Pathology 17, pp. 149-157, 1987.

Zimmermann, R. and Günther, K.P., "Laser-induced chlorophyll-a fluorescence of terrestrial plants", Internat. Geoscience and Remote Sensing Symposium, IGARSS Zürich, Vol. 2, pp. 1609-1613, ESA Publications Division, Noordwijk 1986.

Sensitivity of Reflected Radiation to Organic
Matter Content of Typical Ontario Soils.

A. Nurwadjedi and R. Protz
Land Resource Science
University of Guelph
Guelph, Ontario
N1G 2W1

J. Cihlar
Canada Centre for Remote Sensing
Energy Mines and Resources
Ottawa, Ontario
K1A 0Y7

The future utility of remotely sensed data for measuring soil organic matter content needs to be evaluated for possible use in soil conservation monitoring programs. The basis for this evaluation will be an accurate measurement of the smallest differences in soil organic matter that can be detected using laboratory reflectance systems. In this paper, we examine the sensitivity of reflected radiation to organic matter content of typical southern Ontario soils. Reflectance spectra (470-2450nm) were collected from 174 air dried soil samples with a REFSPEC - 11^A spectrometer. The samples were collected from four transects representing different soil series.

The soil organic matter and clay content of the Perth Silt Loam and Fox Sandy Loam were negatively correlated with all reflectance values over the 470-2450nm range. The red reflectance values were more closely correlated to the organic carbon contents of all soils than were the blue, green and infrared. Three organic matter classes are separable by the red reflectance on the Perth silt loam and the Bookton sandy loam even though the range of organic carbon content is only 3.4 % and 3.2% respectively. The red reflectance could be used to separate two classes of organic carbon in the Huron clay loam (2.40%C) and the Fox sandy loam (6.40%C).

SPECTRAL COMPONENTS ANALYSIS EQUATIONS AND THEIR APPLICATION TO RICE

C. I. Wiegand, M. Shibayama* and Y. Yamagata*

Agricultural Research Service, USDA, Weslaco, TX 78596-0267;

*National Institute of Agro-Environmental Sciences,
Kannondai 3-1-1, Tsukuba, Ibaraki 305, Japan

ABSTRACT

Remote spectral observations expressed as vegetation indices (VI) provide information about the photosynthetic size and net assimilate in plant canopies that permit inferences about crop growth and economic yield. In this paper, the interrelationships are expressed in equation form and are applied to data from an experiment conducted in 1987 at Tsukuba, Japan, with Japonica type paddy rice (*Oryza sativa* L.). The vegetation indices used were the perpendicular (PVI) and normalized difference (NDVI). The equations were validated by the data. Functional relations were predominantly linear. Seasonal cumulative NDVI (Σ NDVI) was more closely related to cumulative seasonal photosynthetically active light absorption (Σ Sp, MJ/m²) and to seasonal dry matter increase (Δ DM, g/m²) than was Σ PVI, but Σ PVI related as closely ($r^2=0.86$) to economic yield (YIELD, g/m²) as did Σ NDVI ($r^2=0.85$). The seasonal cumulations equivalent to area under seasonal PVI and NDVI versus time curves ("spectral profiles") were much more closely related to YIELD than were PVI and NDVI averaged for 3 dates surrounding heading. We conclude that the spectral components analysis (SCA) equations permit interpretation of spectral observations of rice that are consistent with its growth and yield in the paddy.

Key words: Rice, vegetation indices, growth, yield, spectral components analysis, conversion efficiency, net assimilate, photosynthesis, leaf area index

INTRODUCTION

Spectral components analysis (Wiegand and Richardson, 1984; 1987) is a system of equations that interrelates spectral vegetation indices (VI); leaf area index (L); fractional photosynthetically active radiation, PAR, absorbed by the canopy (Fp); cumulative daily PAR absorbed during the season (Σ Sp); above-ground dry matter (DM); and economic yield (YIELD). The equations provide a basis for large area yield estimates from Landsat TM and SPOT data and interface with physiological process models of crop growth and yield. The purposes of this paper are to present the equations and illustrate the spectral-agronomic relationships of their terms using data collected in 1987 at Tsukuba, Japan, for 13 treatments of paddy rice that consisted of incomplete combinations of 3 cultivars, 6 nitrogen (N) application rates and 2 transplanting dates.

THEORY

The equation (Wiegand et al. 19)

$$Fp = g(Lz) = g(h(VI)) = f(VI) \quad [1]$$

shows that Fp can be estimated from remotely observed vegetation indices (VI). Commonly

used VI are the normalized difference (NDVI) (Tucker et al., 1979) and the perpendicular vegetation index (PVI) (Richardson and Wiegand, 1977). In Eq. [1]

$$Lz = h(VI) \quad [2]$$

in which

$$Lz = L/\cos Z \quad [3]$$

where Z = the solar zenith angle at the time of the spectral reflectance observations from which VI are calculated. Lz is used because it is more compatible than L with the sun angle-dependent reflectance factor and PAR absorption measurements (Wiegand and Richardson, 1987).

Fractional PAR (400-700nm) absorbed (Fp) by rice canopies can be expressed (Wiegand et al., 19) by

$$Fp = 0.95[0.96 - 0.06 \exp(-0.5(Lz)) - 0.9 \exp(-0.45(Lz))] \quad [4]$$

Daily absorbed PAR, Sp (MJ/day) is defined by

$$Sp = S * 0.48 * Fp = Sp' * Fp \quad [5]$$

where S = daily incident shortwave

and (300-2800nm) radiation (MJ/day),
 0.48 = the fraction of daily
 shortwave radiation at Tsukuba,
 Japan, that is PAR (Group I, 1985).
 The equation

$$YIELD = q(r(s(LVI))) = p(LVI) \quad [6]$$

in which $YIELD = q(DM2-DM1)$, $DM2-DM1=r(LSp)$ and $LSp = s(LVI)$ explains why grain yield ($YIELD$, g/m² or kg/ha) should be estimable from the area under seasonal VI versus time curves or "spectral profiles". Eq. [6] can be implemented for any seasonal interval of interest; however, economic yield is usually determined at harvest. Thus $DM2$ is aboveground dry matter (kg/m²) at harvest and $DM1$ is aboveground dry matter at transplanting (kg/m²), a number so small that it can be ignored. Thus in Eq. [6] the slopes of the functional relationships designated by q , r , s and p are, respectively, the harvest index, the efficiency of conversion of PAR to dry matter, the efficiency of absorption in terms of photosynthetic size of the canopies expressed by VI, and the yield efficiency in terms of photosynthetic size of the canopy. Eq. [6] is most appropriate for intensive ground studies where weekly or 10-day interval measurements of DM , VI , and L provide the seasonal patterns and smoothing algorithms produce daily estimates of these variables to use in the cumulations. Daily Fp from Eq. [4] times daily incident PAR flux (Sp , MJ/day) defines Sp in Eq. [5] that is cumulated in Eq. [6].

For large area applications, simpler equations are needed. If observations are acquired for cereals between late vegetative development (stem elongation) and mid reproductive (milky grain) development stages when L is at its maximum for the season and relatively unchanging,

$$YIELD=k(Lz) = k(h(VI)) = j(VI) \quad [7]$$

is applicable. The $YIELD = j(VI)$ functional relation can be established for production areas of interest from current ground or satellite spectral observations and $YIELD$ samples from representative commercial fields.

If observations are known to have been acquired at heading (h), then the equation

$$YIELD = n(DMh) = n(o(VIh)) = m(VIh) \quad [8]$$

has utility because for cereals $YIELD$ is usually proportional to DMh . Thus if a vegetation index exists that can estimate DMh satisfactorily, a separate estimate of $YIELD$ from DMh can augment the one directly from VI .

EXPERIMENTAL PROCEDURES

Three cultivars (Koshihikari, Nipponbare, and Shinanomochi) of lowland rice (*Oryza sativa* L.), Japonica type, were grown in concrete-lined paddies in rows 0.25m apart in hills of three plants spaced 0.18m apart. Transplanting dates were 21 May and

June, 1987 (day of year, DOY 141 and 162, respectively). Nitrogen was applied at rates of 0, 2, 4, 6, 8 or 12 g/m² for the plots. Incomplete combinations of treatments resulted in a total of 13 combinations of cultivar-planting date-fertility treatments.

Aboveground dry phytomass (DM) was measured weekly (19 dates) and leaf area index was measured on 7 dates beginning with the fifth date of DM observations. Grain yield ($YIELD$) was determined by harvesting 36 hills at each of 4 sites within each treatment.

Daily shortwave radiation, S (MJ/day) was measured by the Division of Agrometeorology, NIAES, Tsukuba. Fp was not directly measured. The equation of Horie and Sakuratani (1985) was modified to express Fp in terms of PAR and Lz as given by Eq. [4].

Bidirectional reflectance factors were measured on 10 dates during the season using a spectroradiometer (Shibayama et al., 1988) that had a 15° field of view. Measurements were made from 2.5m above the rice canopies at 5nm intervals over the wavelength range 400 to 900nm. The bidirectional reflectance factors (%) for narrow bands centered on 840nm and 660nm were used; they are designated $R840$ and $R660$ in the equations that define PVI and $NDVI$.

The weekly DM and vegetation indices, and the biweekly L measurements were used in third to fifth degree polynomial equations of the form

$$Y = \exp(A_0 + A_1t + A_2t^2 + A_3t^3 \dots) \quad [9]$$

in which $t = DOY/200$; A_0 , A_1 , A_2 , A_3 , and A_4 are the coefficients; and Y designates the respective dependent variable, DM , L , PVI or $NDVI$. There were separate equations for each of the four variables for each of the 13 treatments. These equations provided daily values of DM , L , PVI and $NDVI$ for the cumulations of Eq. [6].

Experimental procedures are reported in more detail by Shibayama et al., 1988; 19.

RESULTS

In the results we will emphasize the comparison of functional relationships in the various equations and equation terms for $NDVI$ versus those for PVI . For 73 observations up to heading, the functional relation of Eq. [2] was

$$Lz = 0.022 PVI^{1.7} \quad [2a]$$

($n = 73$, $r^2 = 0.94$)

where $PVI = 0.778 (R840) - 0.628 (R660) - 1.35$ (Shibayama et al., 1988). For $NDVI$

$$Lz = 0.031 \exp 5.78 (NDVI) \quad [2b]$$

($n = 73$, $r^2 = 0.94$)

where $NDVI = (R840 - R660)/(R840 + R660)$. Since PVI and $NDVI$ each accounted for 94% of the variability in Lz , one can anticipate that they would also be similarly related to other variables as well.

The relation, $Fp = f(VI)$, of eq. [1] for both PVI and $NDVI$ was obtained by inserting

the expressions of Eqs. [2a] and [2b] into Eq. [4] and solving for F_p over the 0 - 38 range of PVI and 0 - 0.96 range of NDVI. These F_p estimates are used, in turn, in Eq. [5] for the ΣSp cumulations of Eq. [6]. The terms in Eq. [6] that contain DM, and the vegetation indices PVI and NDVI are implemented by using the daily estimates of these variables from Eq. [9].

Figure 1 displays three of the terms of Eq. [6] for three separate growing season intervals. In Fig. 1a the $\Sigma Sp = s(\Sigma VI)$ term is shown for PVI and NDVI in separate graphs for each of the three growing season intervals and in Fig. 1b the $\Delta DM = r(\Sigma Sp)$ term is shown for the same three intervals. The three intervals were the first three weeks after transplanting (DOY 141-161 and 162-182, respectively, for early and late plantings), from transplanting to about average date of heading (DOY 141-224 and 162-224), and from transplanting to physiological maturity of the grain (DOY 141-266 and 161-266). The scales on the interior of Figs. 1a and 1b apply to the data for the three week interval right after transplanting. As indicated by the coefficients of determination (r^2) there was more variation in estimating ΣSp and ΔDM from PVI than from NDVI. There is a range among the 13 treatments in ΣSp , ΔDM , ΣPVI and $\Sigma NDVI$ attributable to differences in growth among treatments in response to N fertilization. The first transplanting also achieved a higher L than the second transplanting.

The slopes in Fig. 1a are the efficiency of PAR absorption in terms of the photosynthetic size of the canopies (MJ/m^2 /vegetation index unit). In Fig. 1b, the slopes are the efficiencies of conversion of absorbed PAR to dry matter (kg/MJ in the figures). For NDVI the conversion efficiencies were the same ($2.7 g/MJ$) for the two later growth periods, but for PVI the efficiency was $3.0 g/MJ$ from transplanting to heading and $2.4 g/MJ$ from transplanting to physiological maturity.

Figure 1c shows that in spite of the differences for PVI and NDVI in Figs. 1a and 1b, the seasonal cumulation of PVI and NDVI had about the same coefficient of determination versus economic yield (0.86 and 0.85, respectively). These results show that there is a useful relation between the area under the seasonal vegetation index curve and economic yield. This relation is useful for large area yield estimation where functional relations can be developed for production areas of interest from periodic satellite observations and yields reported by farmers or those obtained directly by sampling a number of fields.

Another relation from Eq. [6] that is of interest (Fig. 2) is

$$\Delta DM = r(s(\Sigma VI)), \quad [10]$$

derived from the $\Delta DM = r(\Sigma Sp)$ and $\Sigma Sp = s(\Sigma VI)$ components of Eq. [6] shown in Figs 1b and 1a, respectively. This equation shows how ΣSp can be eliminated to express the relation directly between plant dry matter changes and cumulative vegetation

indices. The coefficients of determination for data in Fig. 2 were 0.88 and 0.89 for PVI and NDVI, respectively, during the vegetative development phase and 0.80 and 0.94, respectively, from transplanting to physiological maturity of the grain. The $r^2 = 0.94$ between NDVI and ΔDM over the whole season merits comment. Rice canopies remain photosynthetically active right up to harvest as evidenced in this study by green leaf area indices in excess of 1.0 in some treatments at harvest. Even though the DM increase after anthesis is due almost solely to grain enlargement, the close association between DM increase and photosynthesis continues as during vegetative development. This explanation of the high correlation between ΔDM and $\Sigma NDVI$ in Fig. 2b is consistent with the constancy of the efficiency of dry matter conversion ($2.7 g/MJ$) during both vegetative development and the whole season (Fig. 1b) and the high correlation between ΣSp and $\Sigma NDVI$ for the whole season (Fig. 1a).

The vegetation index data averaged for three dates around heading (h) and grain yield at harvest were used in the $YIELD = m(VI_h)$ relation of Eq. [8]. The equations for PVI_h and $NDVI_h$ were

$$YIELD (g/m^2) = 253.8 + 16.9(PVI_h) \quad [8a]$$

$$n=13 \quad r^2=0.63$$

$$YIELD (g/m^2) = 270.4 + 1066.7(NDVI_h) \quad [8b]$$

$$n=13 \quad r^2=0.45$$

These results are much poorer than those in Fig. 1c for the Eq. [6] relation, $YIELD = p(\Sigma VI)$ and indicate the superiority of YIELD estimates from area under the seasonal VI curves over those from VI at one particular time during the growing season. Since the three-date average VI used in this case were obtained between stem elongation and milky grain stages of crop development specified for Eq. [7], these particular results could also be considered the $YIELD = j(VI)$ relation of Eq. [7].

For the data of this study, the $YIELD = n(DM_h)$ relation of Eq. [8] was

$$YIELD (g/m^2) = 254.2 + 435.1(DM_h) \quad [8c]$$

$$n=13 \quad r^2=0.92$$

and the $DM = o(VI_h)$ relations were found to be

$$DM_h = -0.0064 + 0.039(PVI_h) \quad [8d]$$

$$n=13 \quad r^2=0.69$$

and

$$DM_h = -1.09 + 2.31(NDVI_h) \quad [8e]$$

$$n=13 \quad r^2=0.43.$$

Substitution of [8d] and [8e] into [8c] produces the

$$YIELD = n(o(VI_h)) \text{ relations of Eq. [8],}$$

$$YIELD = 251.4 + 16.97(PVI_h) \quad [8f]$$

and

$$YIELD = -220.1 + 1005.1(NDVI_h). \quad [8g]$$

Within the experimental errors associated with [8c], [8d], and [8e], equations [8f] and [8g] predict the same values of YIELD

that Eqs. [8a] and [8b] do, as Eq. [8] requires.

As shown in Fig. 2, DM from transplanting through heading (DOY 224) could be estimated by Eq.[10] as

$$DM(g/m^2) = 74.0 + 0.81(\Sigma PVI) \quad [10a]$$

$$n=13 \quad r^2=0.88$$

$$\text{and } DM(g/m^2) = -.99 + 21.0(\Sigma NDVI) \quad [10b]$$

$$n=13 \quad r^2=0.89.$$

Thus use of PVIh and NDVIh at heading in [10a] and [10b] predicts DMh. Then one could substitute [10a] and [10b] into [8c] to estimate YIELD from DMh.

DISCUSSION

There is considerable flexibility in the use of the equations that constitute spectral components analysis (SCA). One can implement only those parts of the equations of interest or that the data permit. The flexibility is increased if calibrations for terms such as $Fp=f(VI)$ and $Fp=g(Lz)$ of Eq. [1], $DM2-DM1=r(\Sigma Sp)$ of Eq. [6], $YIELD=k(Lz)$ and $YIELD=j(VI)$ of Eq. [7], and $YIELD=n(DMh)$ and $YIELD=m(VIh)$ of Eq. [8] are developed for production areas of interest. The calibrations minimize the number and kinds of observations that need to be made each season.

The SCA equations help establish how inferences about probable yields can be inferred from spectral observations of the canopies. They were devised to provide a basis for large area estimation of crop yields from spectral observations such as those that can be made by orbiting satellites. Thus the relations emphasize spectral vegetation indices and economic yield.

Crop simulation models have been developed that are also capable of estimating economic yields. They require soil property inputs of rooting depth and plant available water, the weather inputs solar radiation, air temperature and precipitation, and reduction factors for applicable stresses. But it is not known for particular fields whether they are actually growing as the model predicts. In contrast, the direct look at the canopies used in spectral components analysis relies on the plants themselves to integrate the growing conditions experienced and to display their responses to them through the canopies that develop (Wiegand and Richardson, 1984). The vegetation indices sense the net assimilate displayed by the canopies as expressed by leaf area index, dry matter and photosynthetic size of the canopies.

REFERENCES

- Group I (Distribution and Utilization of Energy) of the Green Energy Prog. 1985. Bul. No. 4, Seasonal and geographical variation in solar radiation energy. Table II-4-1, p. 60.
- Horie, T. and T. Sakuratani. 1985. Studies on crop weather relationship model in rice.
- (1) Relation between absorbed solar radiation by the crop and dry matter production. J. Agric. Meteorol. 40:331-342.
- Richardson, A.J. and C.L. Wiegand. 1977. Distinguishing vegetation from soil background information. Photogram. Eng. and Rem. Sens. 43:1541-1552.
- Shibayama, M., C. Wiegand, T. Akiyama, and Y. Yamagata. 1988. Radiometric predictions for agronomic variables of rice canopies using a visible to mid-infrared spectroradiometer. Int. Archives Photogramm. and Rem. Sens. 27(7B):508-517. (Proc. 16th Cong. Int. Soc. Photogramm. and Rem. Sens. Kyoto, Japan, July 1-10, 1988.)
- Shibayama, M., C.L. Wiegand, T. Akiyama, and Y. Yamagata. 19 . Spectral observations for estimating the growth and yield of rice. II. Application at Tsukuba, Japan. Jap. J. Crop Sci. (Submitted).
- Tucker, C.J., J.H. Elgin Jr., J.E. McMurtrey III, and C.J. Fan. 1979. Monitoring corn and soybean crop development with hand-held radiometer Spectral data. Rem. Sens. Environ. 8:237-248.
- Wiegand, C.L. and A.J. Richardson. 1984. Leaf area, light interception, and yield estimates from spectral components analysis. Agron. J. 76:543-548.
- Wiegand, C.L. and A.J. Richardson. 1987. Spectral components analysis. Rationale and results for three crops. Int. J. Remote Sens. 8:1011-1032.
- Wiegand, C., M. Shibayama, and Y. Yamagata. 19 . Spectral observations for estimating the growth and yield of rice. I. Constituent equations and their implementation. Jap. J. Crop Sci. (Submitted).

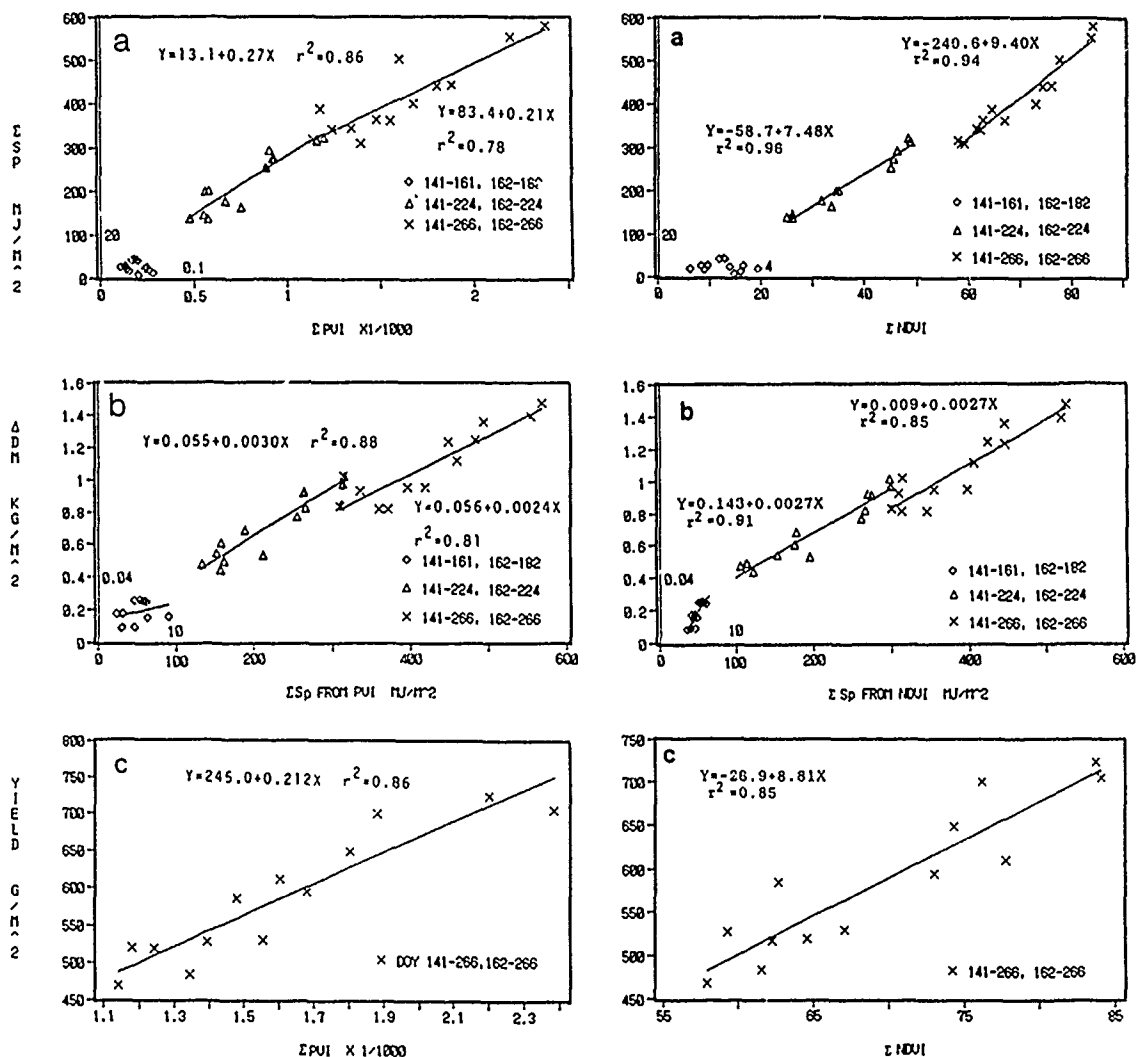


Figure 1. Eq. (6) terms for the 13 treatments of this study expressed using the vegetation indices PVI and NDVI. In parts 'a' and 'b' data are presented for three seasonal intervals: the first three weeks after transplanting (DOY 141-161 and 162-182, respectively, for early and late plantings); from transplanting to approximately average heading date (DOY 141-224 and 162-224), and from transplanting to approximately physiological maturity of the grain (DOY 141-266 and 162-266). The interior scale applies for the three week interval right after transplanting. Part 'a' displays the $\Sigma Sp = s(EVI)$ term of Eq. (6); part 'b' displays the $DM2 - DM1 = r(\Sigma Sp)$ term of Eq. (6); and, part 'c' displays the right side of Eq. (6) for the interval transplanting to physiological maturity.

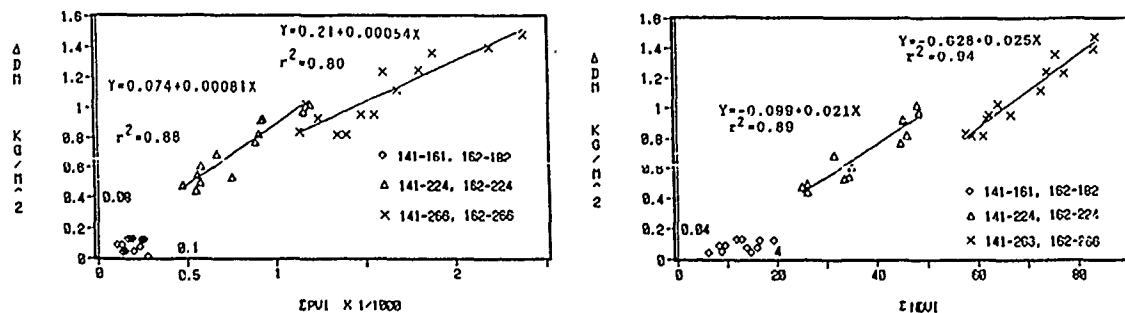


Figure 2. Dry matter changes for the same three seasonal intervals of Fig. 1 versus ΣPVI and $\Sigma NDVI$ as expressed by Eq. (10), $\Delta DM = r(s(VI))$. This relation is the product of the $DM = r(\Sigma Sp)$ and $\Sigma Sp = s(EVI)$ terms shown in Figs. 1b and 1a, respectively.

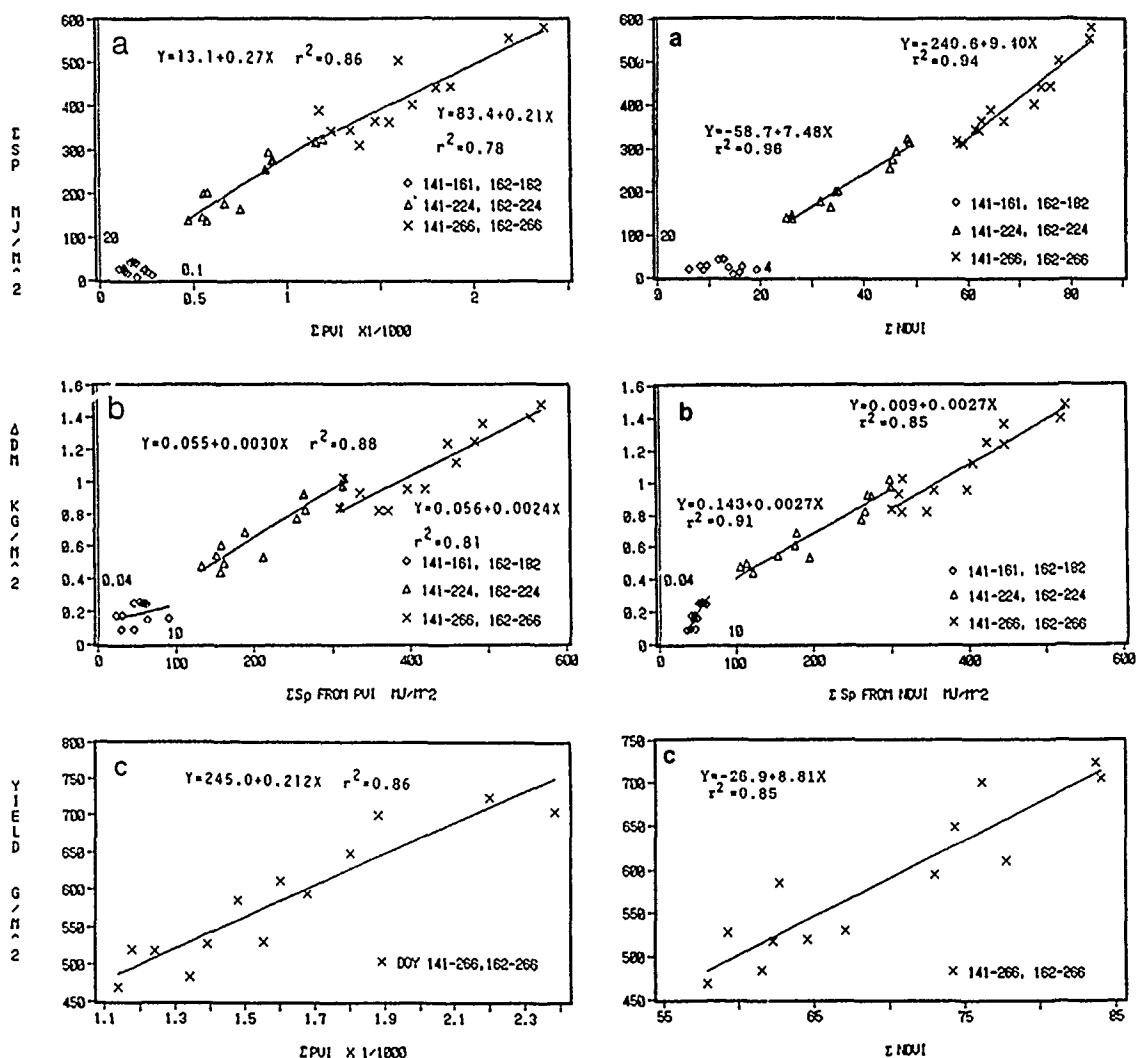


Figure 1. Eq. (6) terms for the 13 treatments of this study expressed using the vegetation indices PVI and NDVI. In parts 'a' and 'b' data are presented for three seasonal intervals: the first three weeks after transplanting (DOY 141-161 and 162-182, respectively, for early and late plantings); from transplanting to approximately average heading date (DOY 141-224 and 162-224), and from transplanting to approximately physiological maturity of the grain (DOY 141-266 and 162-266). The interior scale applies for the three week interval right after transplanting. Part 'a' displays the $\Sigma Sp = s(\Sigma VI)$ term of Eq. (6); part 'b' displays the $DM2 - DM1 = r(\Sigma Sp)$ term of Eq. (6); and, part 'c' displays the right side of Eq. (6) for the interval transplanting to physiological maturity.

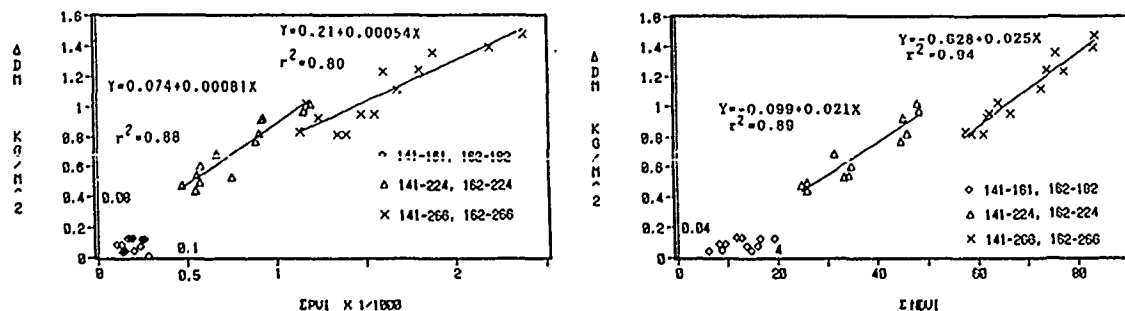


Figure 2. Dry matter changes for the same three seasonal intervals of Fig. 1 versus ΣPVI and $\Sigma NDVI$ as expressed by Eq. (10), $\Delta DM = r(s(VI))$. This relation is the product of the $DM = r(\Sigma Sp)$ and $\Sigma Sp = s(\Sigma VI)$ terms shown in Figs. 1b and 1a, respectively.

UTILISATION DES MESURES RADIOMÉTRIQUES DE TERRAIN POUR LA CARTOGRAPHIE SATELLITAIRE DES SOLS D'UNE RÉGION SEMI-ARIDE DU MAROC

HINSE* Mario, Q.H.J. GWYN*, A. MERZOUK**, F. BONN*

* Centre d'applications et de recherches en télédétection, Université de Sherbrooke, Sherbrooke, P.Qué., Canada, J1K 2R1, tél. (819) 821-7180, télex 05-836149, fax (819) 821-7238

** Institut agronomique et vétérinaire Hassan II, Département des sciences du sol, B.P. 6202, Rabat - Instituts, Maroc, tél. 717-58/59, télex: Agrovét 31873 M, fax 798-85

Mots clefs: Radiométrie de terrain, spectroradiomètre, pédologie, milieu semi-aride, Landsat TM

Résumé

Dans le cadre d'une étude sur la cartographie satellitaire des sols d'une région semi-aride du nord Maroc, une campagne de mesures radiométriques sur le terrain a été réalisée. Le but de cette campagne est de caractériser les types de sols par leurs propriétés spectrales et de démontrer l'effet de paramètres pédologiques sur les réflectances enregistrées. Les mesures de réflectance bidirectionnelle ont été effectuées avec un spectroradiomètre dans la gamme de 400-1100 nm. Les principales classes de sols du secteur d'étude se distinguent par leurs propriétés spectrales. Les bandes spectrales du rouge et de l'infrarouge sont les bandes les plus utiles à discriminer les différentes classes. Des relations évidentes ont été démontrées entre la réflectance et la couleur des sols, le taux de matière organique, le pourcentage d'argile et l'état de la surface des sols.

Abstract

Radiometric studies of the soils in the semi-arid region of Morocco have been made as part of a project to map soils using Landsat images. The purpose of the study was to measure the spectral properties of the soils and compare these with their physico-chemical characteristics. Bidirectional reflectances were measured with a spectroradiometer in the range of 400-1100 nm. The principal soil classes can be distinguished by their spectral signatures. The red and infrared bands are the most useful discriminators. The reflectance is most clearly correlated to the soil colour, organic matter content, clay content and the surface characteristics.

INTRODUCTION

L'application de la télédétection à l'étude des sols contribue à l'identification des types de sols en vue de leur cartographie et elle apporte une information complémentaire dans le but de préciser les caractéristiques pédologiques du sol. Plusieurs auteurs (Cipra et al., 1980; Baumgardner et al., 1985; Mulders and Epema, 1986; Coleman and Montgomery, 1987) ont démontré l'utilité des propriétés spectrales dans l'étude des caractéristiques physiques et chimiques des sols. Ils ont trouvé des relations entre les valeurs de réflectance spectroradiométrique et le taux de matière organique, la teneur en eau, la couleur des sols, le taux de calcaire et la teneur d'oxyde de fer présent dans le sol.

Le but premier de l'étude est de caractériser les types de sols d'une région semi-aride par des valeurs de réflectance acquises sur le terrain, deuxièmement, démontrer l'effet de quelques paramètres physiques et chimiques des sols sur les courbes de réflectance. Cette étude s'inscrit dans le cadre d'un programme sur la

cartographie satellitaire des sols d'une région semi-aride du nord Maroc. Ces mesures radiométriques de terrain représentent une étape préliminaire à la réalisation d'une classification numérique.

SECTEUR D'ÉTUDE

Le secteur d'étude couvre la région de Settat-B'Hammed situé dans la partie septentrionale du Maroc, 75 km au sud de Casablanca. C'est une zone de plaine et de plateaux à vocation agricole de culture céréalière et maraîchère.

Les sols ont été échantillonnés à partir de la carte pédologique au 1 / 100 000. Ils sont compris à l'intérieur de 6 grandes classes de sols (Classification française CPCS et Aubert). Ce sont les classes de sols minéraux bruts, les sols peu évolués, les vertisols, les calcimagnésiques, les isohumiques et les sols à sesquioxyde de fer. Les calcimagnésiques, les vertisols et les isohumiques représentent plus de 70 % de la surface du périmètre d'étude. Ces sols se sont développés à partir de 9 grands types de matériaux roche-mère. Ce sont les marno-calcaire, les argiles rouges, les roches calcaires, les schiste gréseux, les calcaires gréseux encroûtés, les formations superficielles limoneuses rouges et les formations à sables siliceux, les roches calcaires dures et les conglomérats mio-pliocène.

MATÉRIELS ET MÉTHODES

Les données spectrales ont été recueillies lors d'une campagne de terrain du 15 septembre au 15 octobre 1988. L'appareil utilisé est un spectroradiomètre SPECTRON qui mesure une intensité d'énergie lumineuse en fonction de la longueur d'onde ($\mu W/cm^2$). L'instrument est sensible aux longueurs d'onde comprises entre 400 et 1100 nm avec un pas d'échantillonnage de 2,6 nm. L'appareil comprend deux composantes, la première est une tête de détecteurs (photodiodes) avec un angle d'ouverture possible de 1 et 10 degrés. La seconde est un micro-processeur contrôleur qui traite et enregistre le signal.

Toutes les mesures ont été prises de façon perpendiculaire et à une distance de 2 m au-dessus du sol. L'angle d'ouverture de l'appareil était de 10°. La surface prise en compte dans la mesure est d'environ 0,12 m². Dans le but d'éviter un effet rasant de l'éclairement solaire, le temps des mesures était toujours compris entre 9h30 et 15h30. Deux répétitions de mesures ont été réalisées sur chaque site visités. Le facteur de réflectance bidirectionnel (BRF) a été déterminé en procédant à deux mesures consécutives pour chaque cible visée, soit une dans la direction de la surface et l'autre en direction d'une surface de référence. La valeur relative de la réflectance bidirectionnel enregistrée est donnée par :

$$R (\%) = r (t) / s (t) * 100 \quad (1)$$

Où R représente la réflectance de la surface visée, r est la réponse de l'appareil sur le site, t étant le temps d'ouverture d'intégration donné par l'appareil et s la réponse de la surface de référence. Un total de 67 spectres de réflectance ont été recueillis. Le tableau 1 montre la répartition des sites selon leur appartenance aux différentes classes de sols énumérées précédemment.

Toutes les mesures ont été recueillies sur des surfaces de sols nus et secs. Des conditions atmosphériques très bonnes en terme de luminosité et de pureté de l'atmosphère ont permis d'acquies des données stables. La surface des sols a été décrite en fonction de critères qualitatifs tels que la rugosité de surface, les pratiques agricoles appliquées, la pierrosité, la couleur au code de Munsell et d'autres observations à caractères topographiques et géomorphologiques.

Tableau 1

Distribution des sites échantillonnés selon leur appartenance aux différentes classes de sols

| Classes de sols | Nombre de sites |
|---------------------|-----------------|
| Minéraux bruts | 3 |
| Peu Évolus | 2 |
| Calcimagnésiques | 40 |
| Vertisols | 9 |
| Isohumiques | 11 |
| Sesquioxydes de fer | 2 |

De plus toutes les valeurs de réflectance ont été mesurées aux mêmes lieux où furent acquis des échantillons de sols pour une analyse en laboratoire, lors des travaux d'élaboration de la carte pédologique qui se sont déroulés du 11/01/1983 au 16/8/1984. Ceci permettait d'obtenir des éléments physico-chimiques propres aux différents sols en plus des réponses spectrales. Les paramètres de sols retenus pour fin d'évaluation sont la pierrosité, le CaCO_3 total, le taux de matière organique, le pourcentage d'argile, de limon et de sable en plus de la couleur.

RÉSULTATS ET DISCUSSION

Dans le domaine spectral exploré, l'allure générale des courbes de réflectances suit un patron assez similaire d'une de classe de sols à une autre (fig. 1). La réflectance augmente en fonction de la longueur d'onde pour atteindre un maximum vers 900 nm puis décroît graduellement. La forme concave des courbes est marquée sur son parcours par une

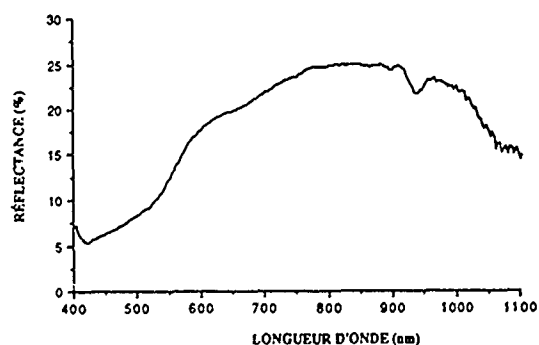


Fig.1. Courbe de réflectance caractéristique des sols de la région de Settati - B'Hammed

dépression généralisée entre 900 et 975 nm. L'amplitude des courbes de réflectance est dépendante des propriétés liées aux états de la surface et des caractéristiques physico-chimiques du sol.

La figure 2 présente les valeurs de réflectance moyenne des six classes majeures de sols. Les données ont été groupées selon cinq bandes spectrales. Les quatre premières correspondent aux bandes TM du satellite Landsat. La cinquième bande regroupe les valeurs de longueurs d'onde complémentaires jusqu'à 1100 nm. Les classes de sols se différencient les unes des autres à l'exception de la classe des calcimagnésiques et des sesquioxydes de fer où l'on remarque une certaine confusion. Les vertisols présentent les valeurs de réflectance les plus faibles et ce pour chaque bande spectrale. La classe de sols des calcimagnésiques, des peu évolués et des minéraux bruts présentent des valeurs de réflectance substantiellement plus élevées que la classe des vertisols et des isohumiques. Ces différences peuvent être attribuées à la nature physique et chimique de ces diverses classes de sols (Tableau 2), à la couleur et aux pratiques culturales. Les bandes rouge et infrarouge se montrent les bandes les plus importantes pour la discrimination entre les classes de sols. Ces résultats corroborent les travaux de Coleman et Montgomery (1987).

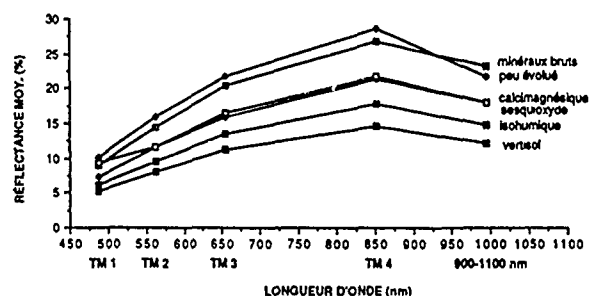


Fig.2 Valeurs des réflectances moyennes du spectroradiomètre, des classes de sols étudiées et groupées en bandes spectrales.

Tableau 2. Valeurs moyennes des propriétés physico-chimiques pour chaque sol.

| Classe de sols | I | II | III | IV | V | VI |
|---------------------|------|------|------|------|------|------|
| Pierrosité (%) | 21,6 | 22,5 | 3,0 | 10,7 | 4,3 | 85,0 |
| CaCO_3 (%) | 3,0 | 0 | 3,3 | 13,6 | 5,6 | 5,0 |
| Mat. Org. (%) | 3,6 | 3,26 | 1,4 | 3,5 | 3,6 | 3,1 |
| Argile (%) | 28,6 | 21,0 | 55,3 | 27,6 | 42,0 | 18,4 |
| Limon (%) | 51,4 | 13,3 | 26,0 | 43,8 | 10,6 | 17,1 |
| Sable (%) | 26,0 | 65,7 | 16,7 | 28,6 | 19,7 | 64,5 |

Classe I = Minéraux bruts II = Peu évolués III = Vertisols
IV = Calcimagnésiques V = Isohumiques VI = Sesquioxydes de fer

Pour les valeurs moyennes de réflectance des classes de sols les plus répandues, un test de comparaison des moyennes (t de student) a été effectué. Les résultats démontrent que les moyennes

des classes de sols pour chaque bande spectrale sont différentes et significatives statistiquement à 95% à l'exception de la classes des vertisols et des isohumiques aux bandes TM1 et TM 2 qui n'ont pas passées le test (Tableau 3).

Tableau 3. Résultats du test t de student pour les valeurs moyennes de réflectance des classes de sols les plus répandues

| Classes de sols | IV - III | IV - V | III - V |
|-----------------|-------------------------------------|------------------------------------|--|
| TM 1 | $T_c=6,23>1,68$ (H_0 rejetée) | $T_c=4,3>1,68$ (H_0 rejetée) | $T_c=-1,2>-1,86$ (H_0 acceptée) |
| TM 2 | $T_c=6,7>1,68$ (H_0 rejetée) | $T_c=4,5>1,68$ (H_0 rejetée) | $T_c=-1,36>-1,86$ (H_0 acceptée) |
| TM 3 | $T_c=10,2>1,68$ (H_0 rejetée) | $T_c=6,4>1,68$ (H_0 rejetée) | $T_c=-5,4<-1,86$ (H_0 rejetée) |
| TM 4 | $T_c=8,6>1,68$ (H_0 rejetée) | $T_c=5,9>1,68$ (H_0 rejetée) | $T_c=-1,9<-1,86$ (H_0 rejetée) |
| 900-1100 | $T_c=8, >1,68$ (H_0 rejetée) | $T_c=5,4>1,68$ (H_0 rejetée) | $T_c=-2,65<-1,86$ (H_0 rejetée) |

H_0 = Hypothèse nulle i.e. $\mu_1 = \mu_2$ niveau d'acceptation = 0.05
Classes de sols III = Vertisols IV = Calcimagnésiques V = Isohumiques

L'observation des différents spectres de la classe de sols des calcimagnésiques a révélé beaucoup de variation à l'intérieur de cette même classe (fig. 3). Trois groupes de sols calcimagnésiques se distinguent clairement soit, les rendzines, les bruns calciques et les sols bruns mélanisés. Les bruns calciques mélanisés présentent les valeurs les plus faibles. C'est un groupe de sols de couleur foncée (10YR 3/2), riche en argile, pauvre en calcaire fin, peu pierreux et de famille à formations limoneuses rouges. Le second groupe de sols, les bruns calciques, sont des sols plus clairs que les précédents, de pierrosité moyenne, moins riche en argile et généralement plus riche en calcaire (de famille marno-calcaire). Ces caractéristiques mettent en place des valeurs de réflectance plus élevées. Le dernier groupe, les rendzines, sont des sols généralement de couleur moyennement plus clair (10YR 5/4 - 7.5YR 5/4) que les deux groupes précédents, ils sont également plus pierreux et aussi formés sur matériau marno-calcaire.

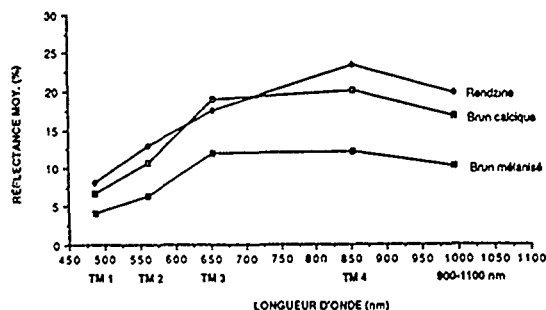


Fig.3. Valeurs de réflectance moyenne de trois groupes de sols de la classe des calcimagnésiques.

En dernier lieu, on a voulu vérifier l'effet de la couleur, de la matière organique, du pourcentage d'argile et des pratiques agricoles sur la réponse spectrale à certains sites (figs. 4, 5, 6, 7).

La couleur d'un sol a toujours été considérée comme l'un des principaux facteurs de discrimination dans les systèmes de classification des sols (Baumgardner, 1985). Selon la figure 4 qui illustre les variations de la réflectance en fonction de la couleur d'une même teinte mais de clarté différente (la couleur des sols est toujours référencé par trois variables: la teinte, la clarté, la pureté) on observe que cette variable est un facteur de variation évidente en relation avec les autres variables pédiques.

En radiométrie de terrain, la matière organique joue également un rôle dans les modifications de la réflectance (King, 1985). On remarque qu'une teneur élevée s'accompagne d'une diminution de la réflectance (Figure 5).

Les classes de sols étudiées montrent des pourcentages d'argile très variés, on a voulu exprimer l'effet de ces taux d'argile sur les réflectances (Figure 6). La réflectance augmente en fonction d'une diminution de la teneur en argile présente dans les sols, et ce pour toutes les longueurs d'onde considérées.

Les pratiques culturales de type labour, hersage ou autres altèrent beaucoup les réponses spectrales enregistrées (Stoner and Horvath, 1971). Ces pratiques interviennent au niveau de la rugosité de la surface. Dans tous les cas observés les réflectances diminuent avec une rugosité croissante (fig. 7).

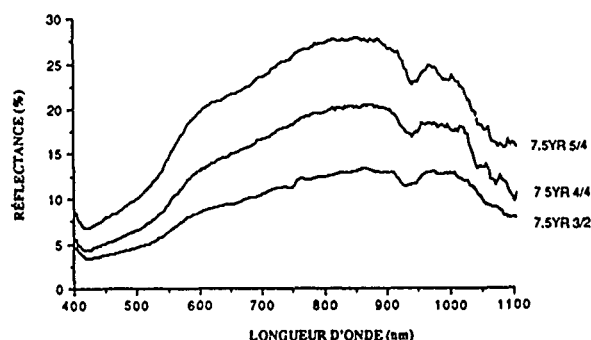


Fig. 4. Influence de la couleur des sols sur les valeurs de réflectance. Sols de même teinte (7.5YR) mais de clarté différente

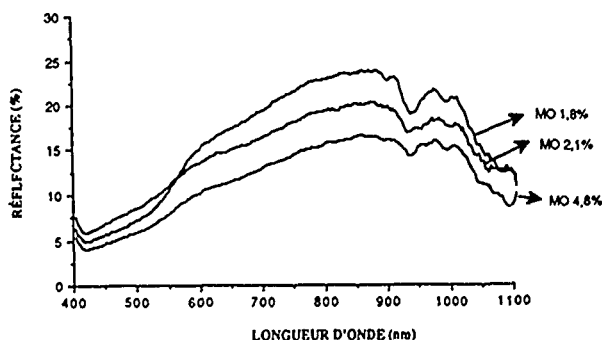


Fig. 5. Effet du taux de matière organique sur les réflectances

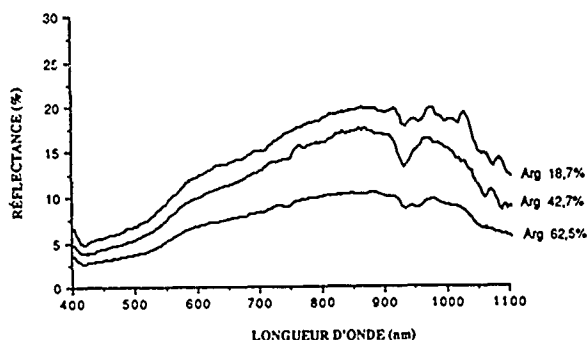


Fig. 6. Effet du pourcentage d'argile contenu dans le sol sur les réflectances

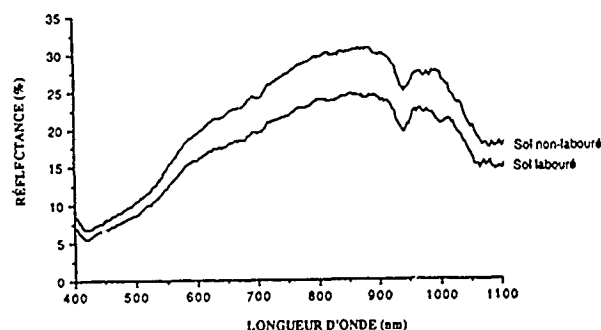


Fig. 7. Influence des pratiques culturales de type labour sur les réflectances d'une parcelle de sol.

CONCLUSION

Ces résultats de mesures radiométriques des sols de la région de Settati - B'Hammed au Maroc permettent de rassembler plusieurs éléments d'information. En premier lieu on a observé qu'il est possible de distinguer spectralement les principales classes de sols rencontrées sur le secteur. Que les bandes spectrales du rouge et de l'infrarouge se montrent les bandes les plus utiles pour la discrimination entre les différentes classes. De plus il est possible de distinguer trois groupes de sols différents à l'intérieur de la classe des calcimagnésiques soit les rendzines, les bruns calciques et les bruns mélanisés.

Ils ont permis de mettre en évidence des relations claires entre les différents paramètres pédologiques et les modifications de la réflectance enregistrée dans la région de 400 à 1100 nm. La réflectance est étroitement liée à la couleur du sol qui à son tour dépend de plusieurs caractères physico-chimiques. Le taux de matière organique et le pourcentage d'argile altèrent les réponses spectrales, la réflectance diminue avec une augmentation de la matière organique et du pourcentage d'argile présents dans les sols. Finalement on observe qu'un sol labouré a toujours une réflectance moindre qu'un sol non labouré.

Pour être plus complète cette étude nécessiterait un nombre d'observations plus élevées pour certaines classes de sols étudiées. Il faut considérer qu'un sol est une cible complexe qui évolue dans le temps et dans l'espace où les interactions des plusieurs paramètres sont liées. Néanmoins les informations qui ressortent de cette étude seront d'un grand recours lors de l'interprétation des résultats obtenus à partir des traitements numériques de l'image satellite.

REMERCIEMENTS

Nous remercions Abdelatif Filali et Lahoucine Serrau du Département des sciences du sol de l'Institut Agronomique et Vétérinaire Hassan II pour leur participation aux travaux de terrain. Le Centre de recherche en développement international (CRDI) du Canada a subventionné ce projet (Contrat de recherche 3-P-87-1023).

RÉFÉRENCES

- Baumgardner, M.F., L.F., Silva, L.L., Biehl and E.R., Stoner, "Reflectance properties of soils", Academic Press Inc., New-York, *Advances in Agronomy*, Vol. 38, pp.1-44, 1985.
- Cipra, J.E., D.P., Franzmeier, M.E. Bauer and R.K. Boyd, "Comparison of multispectral measurements from some nonvegetated soils using Landsat digital data and a spectroradiometer", *Soil Science Soc. Am. Journal*, Vol. 44, pp. 80-84, 1980.
- Coleman, T.L. and O.L., Montgomery, "Soil Moisture, organic matter, and iron content effect on the spectral characteristics of selected vertisols and alfisols in Alabama", *Photographic Engineering and Remote Sensing*, Vol. 53, No.12, pp.1659-1663, 1987.
- Girard, M.C., "Télédétection de la surface du sol", Applications de la télédétection à l'agriculture, Séminaire, Proc., Paris, France, pp. 177-193, 1983.
- King, C., "Études des sols et des formations superficielles par télédétection ; approche de leurs caractéristiques spectrales, spatiales et temporelles dans le visible et le proche infrarouge", Document du B.R.G.M. - No 96, Thèse soutenue à l'Institut national agronomique, Paris - Grignon, 174 pa., 1985.
- Mulders, M.A. and G.F., Epema, "The thematic mapper : a new tool for soil mapping in arid areas", *ITC Journal, Proc. Symp. ISSS*, pp. 24-29, 1986-1.
- Stoner, E.R., and M.F., Baumgardner, "Characteristic variations in reflectance of surface soils", *Soil Science Am. Journal*, Vol. 45, pp. 1161-1165, 1981.
- Stoner, E.R., and E.H., Horvath "The effect of cultural practice on multispectral response from surface soil", *Proc. of the 7th Symposium on Remote Sensing of Environment*, Vol. III, Erim, pp. 2109-2113, 1971.

Basis Frequency Analysis, BAFRAN, - towards a fresh approach to analyse the spectrum of a short record.

Michael P. Beddoes

ABSTRACT

The Basic Frequency Analyser, BAFRAN, aims to produce an estimate of signal spectrum from a relatively short signal record or from equal length signal records. The signal is modelled as the sum of infinite extent complex exponentials even though the signal itself is known for only a brief period T . Ideally, the spectrum detected should only include signal components which are multiples of $2\pi/T$; contributions from other signal components are suppressed. The ideal operation has a drawback: a signal component of frequency $2\pi/T$ (part of a "basis" set) will appear at the output of the analyser but a signal component at $3\pi/T$ will be suppressed. A way around this problem is described. A strength of the approach is that spectrum leakage, normally encountered when taking the spectrum, is minimized.

Three methods to realize BAFRAN have been proposed but shortness of space only allows description of one. In this method, multiple records are averaged to suppress non-basis components in the signal. The method is simple to instrument.

Introduction

The area of spectrum estimation is a fascinating one [1-5]. Here we will describe a new estimator of spectrum which is, at the present, looking for an application.

We shall assume that the signal $f(t)$ is given by

$$f(t) = \sum_{k_1=-\infty}^{\infty} a_{k_1} \exp j\omega_{k_1} t + \sum_{k_2=-\infty}^{\infty} a_{k_2} \exp j\omega_{k_2} t \quad (1)$$

$\omega_{k_1} = 2\pi k_1 / T$ where k_1 is integer, $\dots, -2, -1, 0, 1, 2, \dots$, and T is a constant. k_2 is a plus or minus continuous variable which excludes all integers.

The left hand set of terms in (1) comprise what is known as a "basis" set determined by T ; the right hand set of terms form the non-basis set.

The Continuous Fourier transform is given by

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp -j\omega t \, dt \quad (2)$$

This relates a continuous (in ω) spectrum to a continuous (in t) signal. If $f(t) = \exp j\omega_1 t$, $F(\omega) = \delta(\omega_1 - \omega)$. Suppose we restrict the ω variable to basis frequencies only given by $\omega^1 = k \omega_0$, (k integer) where $\omega_0 = 2\pi / T$, then we would get a "sifted" spectrum given, by

$$F(\omega) = \begin{cases} \delta(\omega - \omega^1) & \text{if } \omega^1 = k \omega_0, \\ 0 & \text{for } \omega^1 \neq k \omega_0. \end{cases} \quad (3)$$

The expression (3) is a true projection of the spectrum on the basis set of frequencies. Bafran attempts to realize the transformation of equation (3) using sampled signals.

The finite Fourier Transform is obtained from (2) by altering the limits and taking sampling into account.

$$F(k) = \sum_{n=0}^{N-1} f(n) \exp -j2\pi(k - N/L)n \quad (4)$$

$0 < k < N-1$

Using (4), and the mono-component signal $f(n) = \exp 2\pi n/L$,

$$F(k) = \sum_{n=0}^{N-1} \exp -j2\pi \frac{n}{N} (k - N/L) \quad (5)$$

N/L is the normalized frequency. If N/L is integer, the monocomponent signal is basis and $F(N/L) = N$, $F(k \neq N/L) = 0$. $F(k)$ is a plausible representation of the spectrum. If $N/L \neq$ integer, none of the $F(k) = 0$, k integer, and we lose the precision in the

representation.

The Bafran Method.

Corresponding samples from several contiguous equal length records are averaged.

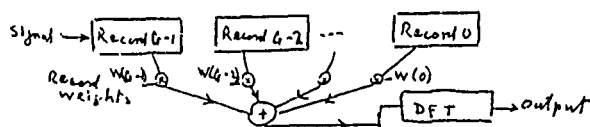


Figure 1:

Block Diagram of apparatus to realize BAFRAN.

Let a typical signal component be

$$s(n) = e^{j(2\pi/L)(n+M) + k}, \quad (6)$$

n is the sampling index which ranges from 0 to $N-1$. N is the number of sampling points per record, and L is the wavelength (in sample points) of the monocomponent. M is a record delay $= gN$.

Then, the sum of the Fourier transforms of G records = Fourier transform of the sum of G records,

$$\begin{aligned} S(k, G) &= \sum_{g=0}^{G-1} \sum_{n=0}^{N-1} \{e^{j(2\pi/L)(n+gN)} e^{-j(2\pi nk)/N}\} \\ &= \sum_{g=0}^{G-1} \{e^{j(2\pi gN)/L}\} \sum_{n=0}^{N-1} \{e^{j(2\pi n)(1/L - k/N)}\} \end{aligned} \quad (7)$$

Term 1 Term 2

If $N/L = \text{integer}, I$, then $S(I, G) = GN$. $S(k/I, G) = 0$. If N/L is not unity, term 2 returns a fixed value N' , where $|N'| < N$. Term 1 will be the sum of unit vectors with a modulus less than G . Explicitly, term 1 is given by

$$\sum_{g=0}^{G-1} e^{j2\pi gN/L} = \exp j \frac{\pi}{L} (G-1)N \frac{\sin G \pi (N/L)}{\sin \pi (N/L)} \quad (8)$$

Figure 2b gives a plot, for $G=51$, of the decibel magnitude of (8) to a base of normalized frequency $\theta = 2\pi n/L$. (8) is cyclic and $\theta = 0$ corresponds to a basis frequency. On either side the central lobe falls to minus infinity. (The plot is limited by practical considerations to show only -36 db). Beyond the lobe, the decibel plot rises to a pair of maxima at -15 db. Worst-case lobe response is -15 db.

Unequal record-weights can be used. Cancellation of non-basis terms can be made more effective. For example, triangular weights give worst case lobe response = -28 db.

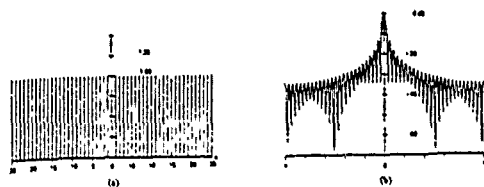


Figure 2 Log magnitude equation 8 for $G=51$, uniform distribution.

2a = weights for each record,
2b = log-magnitude plot.

Fractional signal frequencies.

One can premultiply the $f(n)$ with a suitable delay. e.g. $f'(n) = f(n) \exp -j \frac{\pi}{N} n$. $f'(n)$ when used with Bafran will return an $F'(k')$ where the spectrum will correspond to $k' + \frac{1}{2}$. Other fractional frequency components can be similarly detected.

CONCLUSION

BAFRAN aims to reproduce a spectrum from a short record in which only basis signal terms are included. BAFRAN basically models the signal as the projection on the basis set of frequencies of the spectrum form infinite extent exponentials. Errors in spectrum estimation produced by "spectrum leakage" should be thereby minimized.

A method has been described which is simple to instrument. We programmed the method on a PDP 11; results were plausible. There is a trade-off between the size of the error and the number of records used to determine the spectrum. Other methods could be used to realize the analyser but shortness of space prohibits description.

Acknowledgements.

Initial experiments were conducted on a main-frame computer at Imperial College London, United Kingdom, 1985-86. The author wishes to thank Dr. Richard Kitney and others on the staff at IC for assistance and discussion. A travel grant from NSERC (Canada) is acknowledged with gratitude.

REFERENCES

1. Robinson, E.E., "A Historical Perspective of Spectrum Estimation", Proc. IEEE, Vol. 70, 1982, pp 885-907, especially pages 902 and 903.

2. Burg, J.P., "Maximum Entropy Spectral Analysis", Proceedings of the 37th Meeting of the Society of Exploration Geophysicists, 1967.
Reprinted in IEEE Press, "Modern Spectrum Analysis" (ed. Childers, D.G.) 1978, pp 34-48.
3. Burg, J.P., "A new analysis technique for the time series data" presented at the NATO Advanced Study Institute on Signal Processing on Underwater Acoustics, Aug. 12-23, 1968. *ibid.* pp. 42-48.
4. Ulrych, T.J., and Bishop, T.N., "Maximum Entropy Spectral Analysis and Autoregressive Decomposition", *Rev. Geophysics and Space Physics*, Vol. 13, pp 183-200, Feb. 1975.
A compendious applications paper of MEM with a copious review.
5. Harris, F.J., "On the use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proc. IEEE*, Vol. 66, 1978, pp. 51-83.

CONVERGENT FRONTS IN A PARTIALLY MIXED ESTUARY

Anderson, J.M., Evans, H. and Djavadi, D.

Department of Applied Physics and Electronic
& Manufacturing Engineering
University of Dundee,
DUNDEE DD1 4HM, Scotland, U.K.

Abstract

The use of airborne MSS data in contributing to the understanding of the frontal systems and complex circulation patterns in the estuary of the River Tay in Scotland is described. Particular emphasis is placed on the importance of the thermal infrared data in the detection of frontal systems between water masses of different salinity. The presence of lines of foam or debris visible at particular states of 'he tide is described and the likely mechanisms for the formation of these foam lines are discussed.

1. Introduction

A striking feature often observed in estuaries is the presence of lines of foam or debris often (but not always) aligned with the flow patterns within the waters of the estuary or with the prevailing wind. It is the purpose of this paper to review some of the current knowledge concerning the nature of these foam lines and to discuss in particular the light which study of them throws on the very complicated nature of the circulation of discrete water masses of varying salinity in the estuary of the River Tay in Scotland.

2. Study Area

The estuary or firth of the River Tay stretches for approximately 50 km from the City of Perth to the North Sea. It is navigable for ships of up to 20000 tons as far as the City of Dundee and by coasters of up to 2000 tons to the harbour at Perth. The catchment area of the river extends to 6500 square kilometers including some of the highest ground in the U.K. and this makes it easily the river of greatest fresh water flow in Great Britain. Tidal currents in the estuary can be up to 5 or 6 knots during springs, the mean tidal range being from 4.7 m (springs) to 2.5 m (neaps). The Firth of Tay exhibits a wide variety of water types ranging from pure fresh low suspended sediment concentration (SSC) water in the upper reaches of the estuary through brackish high SSC water in the middle reaches to salt North Sea water of relatively low SSC at the mouth. The marine water flowing into the estuary has a salinity of 35 g l⁻¹ compared to the negligible salinity of the fresh water flow. The freshwater flow varies considerably over the year. The average flow rate is 198 m³/s but the maximum can rise to 1550 m³/s during winter and early spring and the minimum can be as low as 15 m³/s

during drought conditions in May, June and July. Mixing of the water masses takes place but the topography of the estuary bed and extensive sandbank system can interfere with the mixing process resulting in the creation of discrete water masses of varying salinity throughout the estuary especially in the middle reaches. The highest SSCs are observed in the central regions of the estuary where concentrations can be as high as 200 mg l⁻¹ (or 400 mg l⁻¹ at the turbidity maximum) compared to the fresh water high of 10 mg l⁻¹ and the salt water norm of 10 mg l⁻¹ (with an exceptional 50 mg l⁻¹ after storm conditions). Figure 1 shows the form and topography of the estuary. This paper will be confined to discussion of the frontal systems in the area shown since this represents the area of greatest activity. The main shipping channel follows the southern shore line as far as Balmerino, then swings northwards and broadens towards the Rail Bridge. A second channel drains the northern tidal flats and continues past the Road Bridge before joining up with the main channel off Newcome Shoal.

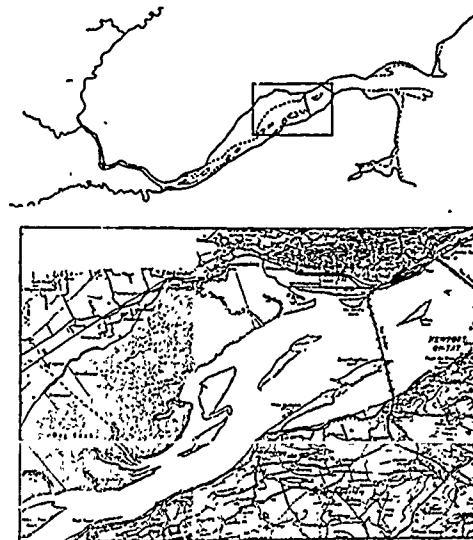


FIGURE 1. The area studied is indicated. It covers the middle reaches of the estuary and stretches from the Road Bridge to Flick Point.

3. Data Acquisition

The study area has been included in the NERC (National Environmental Research Council) Airborne Campaign each year since 1985. This Campaign uses a Piper Chieftain aircraft fitted with a Daedalus AADS 1268 scanner flown during spring, summer and autumn periods. Weather conditions have not been too kind in this area and so far the most successful flights over the Tay have been those flown during the autumn of 1985 and the spring of 1986. 1988 data extending the area of coverage while of good quality is only available for low tidal conditions and is not relevant to this particular study. Concurrently with the remotely sensed data acquisition water sampling from two boats was carried out. Near surface temperature, salinity and SSC was determined throughout the area so far as possible coinciding with each aircraft sortie. Three sorties were undertaken each year covering an ebb tide in 1986 and a flood tide in 1985. Low cloud meant that data for both 1985 and 1986 was gathered from a height of 600 m. This has the advantage that a spacial resolution of 1 m is achieved but the disadvantage that many passes are needed to cover the area with a resultant very large amount of data processing.

4. Water-truth Data

Figure 2 illustrates the way in which the salinity

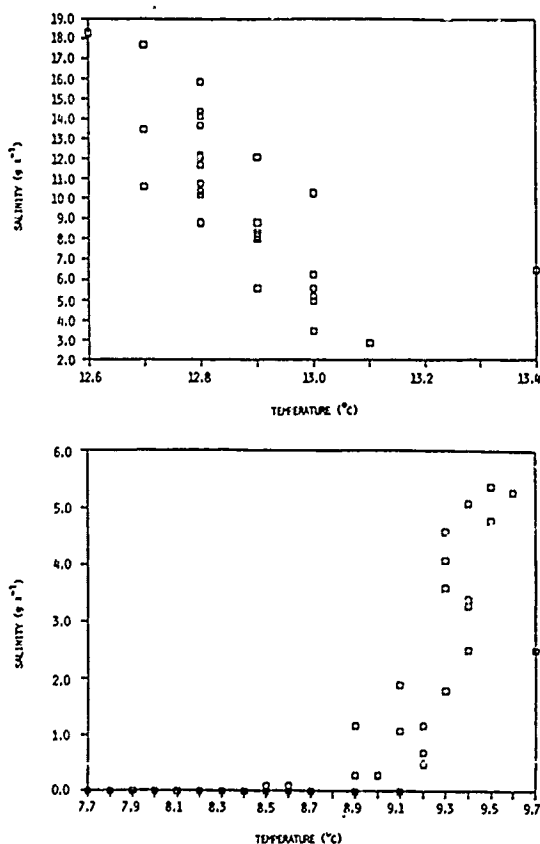


FIGURE 2. The upper graph indicates measurements taken in autumn 1985 and the lower measurements taken in the spring of 1986.

varies with temperature for each of the years 1985 and 1986. Each point represents measurements taken at a particular station chosen randomly in the study area. It can be seen that during the autumn period in 1985 the saltier the water the colder it tended to be whereas during the spring period in 1986 the salt water tended to be considerably warmer than the fresh or blackish water. Figure 3 shows plots of SSC against salinity for the same periods. Although the 1986 data was taken on an ebb tide and the salinity throughout the area was much less than for the flood conditions of 1985 it can still be seen in both cases that the highest SSC's are evident for the water in the middle of the salinity range. This corresponds to the water in the middle reaches of the estuary which has been subjected to turbulent mixing and wave action in the vicinity of the sand banks and mud flats.

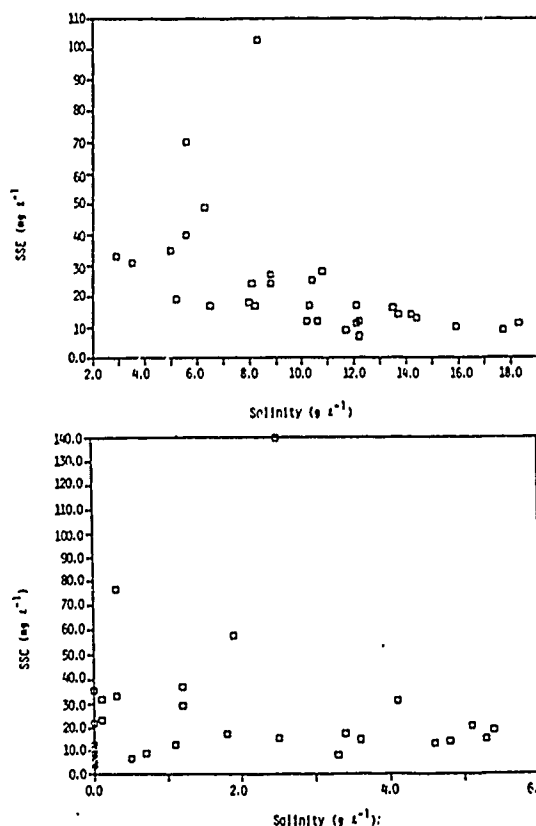


FIGURE 3. The upper graph shows the way in which SSC varies with salinity in autumn 1985 and the lower the variation in spring 1986.

5. Aircraft Data

The positions of the frontal systems marked by lines of foam in the study area is illustrated for a mid-state of the tide in each year in figure 4. It should be noted that the picture changes quite rapidly throughout the tidal cycle but that in general the areas of greatest foam line creation activity are associated with channels in the vicinity of sandbanks and that flood tide activity is greater than ebb tide activity. The numbered fronts will be referred to in the discussion. The area between the bridges was not covered in 1986 and the area up-estuary beyond Balmerino was not covered in 1985.

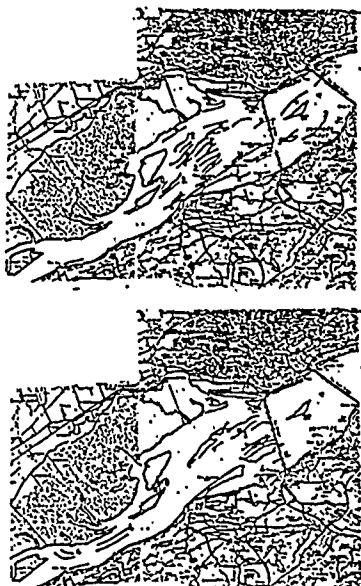


FIGURE 4. The position of foam lines are shown for a flood tide (upper) in autumn 1985 and for a spring tide (lower) in spring 1986 two hours from high tide.

6. Foam Line Creation Mechanisms

Foam lines will be created by the bringing together of suspended impurities where surface waters converge. Such convergence can be created by a variety of mechanisms including:-

(1) Axial convergence where the driving mechanism is the lateral density gradient caused by the intruding tidal salt water during flood tide conditions.

A two-celled, full-depth secondary flow mechanism has been proposed by Nunes and Simpson which results in surface convergence, down welling in mid-channel and divergence at the bed in relatively narrow channels. Such foam lines would not have associated with them a change of salinity, SSC and temperature due to the presence of discrete water masses on either side of the foam line and would possibly only be present during the 1985 tidal conditions.

(2) Confluence of water masses of differing density caused by flow conditions near sandbanks or at the edge of the principal channels.

A variety of mechanisms could be proposed here but the basic principle is that the saltier, denser water will sink and attempt to flow under the less dense fresh water causing a lowering of the surface at the front and resultant convergence. It would be expected that the resultant foam line would have a component of velocity at right angles to the front but the flows associated with such density gradients are considerably smaller than the tidal currents producing the confluence and the line of foam would be expected to extend along the interface of the front in the direction of the current. Such foam lines would have associated with them changes in salinity, SSC and temperature across them.

(3) Foam lines which have become separated from the mechanism which produced them.

During windy conditions a line of foam can be wind-stabilised and become detached from the area where it is originated. Such a line of foam would be aligned with the wind and would not necessarily be associated with any change in surface conditions in its immediate vicinity.

(4) Langmuir circulation.

This is a wind induced double circulation mechanism which can lead to surface convergence. Lines of foam associated with them are regularly observed aligned with the wind direction in the waters of lochs when the wind strength exceeds 3-4 m/s. Such foam lines would not be associated with a frontal system and are usually associated with fairly still waters.

7. Discussion

It has been shown by Rimmer, Collins and Pattiaratchi that airborne thematic mapper data can be used to quantitatively detect changes in salinity as well as changes in SSC. The complicated nature and rate of change of the flow patterns and shallowness of the water outside the channels in this estuary however make direct comparison of water-truth data over the whole scene very difficult. With only two boats available it is not possible in fact to gather a data set large enough in the short time available to enable direct correlation with the remotely sensed data to be carried out. An alternative method for detecting the frontal systems between water masses of differing quality was therefore used. The water-truth measurements illustrated earlier give the picture of a partially mixed estuary which has low SSC fresh water in the upper reaches, low SSC salt water near the outer reaches and high SSC brackish discrete water masses in the middle reaches. In the autumn the intruding salt water is warmer than the injected fresh water and the reverse is true in the spring. The brackish water masses will in each case be at an intermediate temperature. Some complication will arise due to heating of water flowing over solar warmed sand and mud flats but it is not thought that this effect is a large one due to the high flow rates and large quantities of water in the channel systems. These observations show why frontal systems are apparent in the MSS data. A front between two water masses of differing salinity will show up in the thermal infrared channel (11) with the fresher of the two water masses appearing as the high temperature side of the front in the autumn and as the colder side of the front in the spring. When viewed in the visible channels the side of the front with highest SSC will be brackish water and the water on the side with lower SSC will be either fresh (if warmer in the autumn or colder in the spring) or salt (if colder in the autumn and warmer in the spring). In this way by examining the images produced by the data the characteristics of the water masses on either side of a frontal system can be inferred.

Some of the foam lines indicated in Figure 4 are indicated as being attached to a frontal system by the letter "H". This indicates the high temperature side of the front and allows the identification of the position of the saltier water mass. Only those temperature differences are marked where a significant change occurs over the front. The changes range from 0.1°C to 3.0°C. In these cases it is suggested that the foam line is being formed at the confluence of two water masses of differing salinity. The proximity of sandbanks in each case tends to confirm this view.

Figure 5 shows an image taken from the 1986 data which illustrates such a confluence quite clearly. In this



FIGURE 5. A foam line formed on the confluence of two water masses.

case the warmer and therefore saltier water is indicated by the letter "S". In order that such a formation of a foam line could be checked an effort was made in the summer of 1987 to detect a frontal system and predict the position of the formation of a foam line. At this period of the year (July) the fresh water is warmer than the sea water and the salinity and temperature were measured during traverses in the vicinity of the foam line marked "1" in Figure 4 one hour after high water when no evidence of foam line production was visible on the surface. Eventually the results shown in Figure 6 were obtained and within a few minutes the foam started to gather on the surface as illustrated in Figure 7. Measurements of salinity with depth on both sides of the front showed an increase with depth.

The situation can be seen to be quite complicated with several foam lines appearing in the data which do not appear to be connected with frontal systems. It is certainly the case that in places in the data a line of foam can be seen to be wind-dissociated from a frontal system but in general a certain amount of disruption of the line is visible. The greater frequency of the foam lines in the 1985 flood-tide data is suggestive of the likely occurrence of axially convergent fronts forming at this state of the tide. The position of such a system is likely marked by the position of the foam lines marked 2 in mid channel near the West Naughton bank since these do not appear during the ebb-tide. Similarly the foam line marked 3 off Newport is a likely candidate. It is not impossible however that an axially convergent system could happen on an ebb tide under the right conditions. McManus has shown that it is possible for a mass of salt water to penetrate up-estuary, traverse

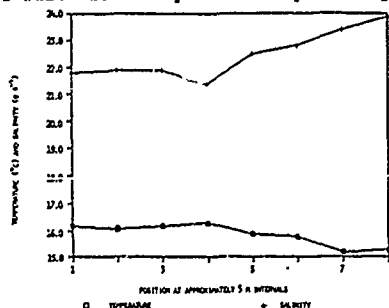


FIGURE 6. Results taken during a traverse across a front are shown.



FIGURE 7. A foam line forming on the front illustrated by Figure 6.

the estuary and return seaward on the ebb. Such a mechanism cannot at present be ruled out in the case of fronts 4 since increases in salinity have been observed during ebb tide conditions at Newport Pier.

Given the high flow rates in the channels where most of the foam line production takes place and the absence of any foam lines in the data so far obtained in the upper reaches of the estuary where little or no salt water intrusion takes place it is not possible to identify the existence of Langmuir circulation cells in the estuary from the MSS data. In fact given the scarcity of foam lines during slack water conditions this can probably be rejected as a major mechanism within the estuary for the present.

8. Conclusions.

It has been demonstrated that the channel 11 MSS data in particular can be used to demonstrate the presence of frontal systems associated with the foam lines detected in the visible data. Two types of mechanisms are proposed as being the main contributors to foam line production - axially convergent systems present when a mass of salt water intrudes into a channel of fresher water (principally on the flood tide but exceptionally on the ebb tide) and confluence systems where two water masses of different salinity are brought together during flood or ebb conditions, the controlling mechanism being the flow pattern generated by sandbanks or bottom topography.

9. References.

1. Nunes, R.A. and Simpson, J.H., "Axial Convergence in a Well-mixed Estuary", *Coastal and Shelf Science*, Vol. 20, pp 637-649, 1985.
2. Rimmer, J.C., Collins, M.B. and Pattiaratchi C.B., "Mapping of Water Quality in Coastal Waters using Airborne Thematic Mapper Data", *International Journal of Remote Sensing*, Vol. 8, No. 1, pp 85-102, 1987.
3. McManus, J. and Wakefield, P., "Lateral Transfer of Water across the Middle Reaches of the Tay Estuary", *Research Report No. 7, Tay Estuary Research Centre*, 1982.

Acknowledgements

The authors would like to thank the staff of the NERC and of the Tay Estuary Research Centre for their help in the data acquisition necessary to this study.

CLASSIFICATION OF PUGET SOUND NEARSHORE HABITATS
USING AIRCRAFT MULTISPECTRAL SCANNER IMAGERY

by

Frank Mynar, II, Ridgeway D. Weerackoon and Mark V. Olsen
Lockheed Engineering and Sciences Company
Spatial Analysis Laboratory
1050 E. Flamingo Rd., Suite 120
Las Vegas, NV 89119
(702) 789-3151

and

Ross S. Lunetta
U.S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
P.O. Box 93478
Las Vegas, NV 89193
(702) 798-2175

ABSTRACT

As part of a comprehensive estuarine management program being administered by the U.S. Environmental Protection Agency (EPA), the nearshore habitats of the Puget Sound area are being inventoried through the use of remote sensing technologies. This inventory, performed at the EPA's Environmental Monitoring Systems Laboratory in Las Vegas, Nevada, utilizes aircraft-borne multispectral scanner (MSS) data to classify marine and estuarine habitats of the intertidal zone.

Classification protocols developed at a test site, the Dungeness Bay area, are detailed. A variety of pre-classification and classification techniques are investigated at this site, for purposes of identifying an "optimal" set of procedures. These techniques include principal components analysis, masking of raw data to eliminate extraneous information, classifications using combinations of principal components output channels and raw data, as well as supervised, unsupervised, and hybrid classification approaches. Additionally, the feasibility, procedures, and costs of transferring the results of the classification to a GIS system, for storage, display, and further analysis of the habitat information, are discussed.

Protocols developed at the Dungeness site will be extrapolated to additional test sites for further refinement, and, ultimately, will be applied to over a large portion of the Puget sound area.

USE OF THEMATIC MAPPER DATA FOR MAPPING OF WATER QUALITY¹KAI SØRENSEN¹, JAN NILSEN², HANS VIGGO SÆBØ³ AND ERIK HOLBÆK-HANSEN³.¹Norwegian Institute for Water Research, Blindern, N-0313 Oslo, Norway²Norwegian Hydrotechnical Laboratory, N-7034 Trondheim, Norway.³Norwegian Computing Centre, Blindern, N-0314 Oslo, Norway.

ABSTRACT

The use of Landsat Thematic Mapper data for mapping and monitoring of water quality in the Oslofjord area in southern Norway have been tested. The test area comprises a mixture of polluted fresh water with different optical properties and sea water which sometimes has high algal concentrations. In situ optical as well as traditional water quality parameters have been measured and compared with satellite data. Different models and water classifications methods have been tested. Some of the results and conclusions of the possibilities of using satellite data in pollution monitoring programs is presented.

Models expressing the relationship between water quality parameters and satellite data have been established. Mainly linear models have been tested and significant correlations were found for the light beam attenuation coefficients, secchi depth, turbidity and suspended sediments. The in situ optical parameters did not correlate with satellite significantly better than the water quality parameters. TM band 1 and 3 usually contributed most to explain these parameters, but the other bands could also contribute to the models. Classification methods gave no new information compared with the modelling. The distribution of the polluted river water are easily seen and correlated with the satellite data. Algal blooms where sometimes difficult to interpret, but during a summer situation interesting information of algal patchiness at the boundary to the nutrient-rich river water where detected.

USE OF THEMATIC MAPPER DATA FOR MAPPING OF WATER QUALITY¹KAI SØRENSEN¹, JAN NILSEN², HANS VIGGO SÆBØ³ AND ERIK HOLBÆK-HANSEN³.¹Norwegian Institute for Water Research, Blindern, N-0313 Oslo, Norway²Norwegian Hydrotechnical Laboratory, N-7034 Trondheim, Norway.³Norwegian Computing Centre, Blindern, N-0314 Oslo, Norway.

ABSTRACT

The use of Landsat Thematic Mapper data for mapping and monitoring of water quality in the Oslofjord area in southern Norway have been tested. The test area comprises a mixture of polluted fresh water with different optical properties and sea water which sometimes has high algal concentrations. In situ optical as well as traditional water quality parameters have been measured and compared with satellite data. Different models and water classifications methods have been tested. Some of the results and conclusions of the possibilities of using satellite data in pollution monitoring programs is presented.

Models expressing the relationship between water quality parameters and satellite data have been established. Mainly linear models have been tested and significant correlations were found for the light beam attenuation coefficients, secchi depth, turbidity and suspended sediments. The in situ optical parameters did not correlate with satellite significantly better than the water quality parameters. TM band 1 and 3 usually contributed most to explain these parameters, but the other bands could also contribute to the models. Classification methods gave no new information compared with the modelling. The distribution of the polluted river water are easily seen and correlated with the satellite data. Algal blooms where sometimes difficult to interpret, but during a summer situation interesting information of algal patchiness at the boundary to the nutrient-rich river water where detected.

A DATA-FUSION APPROACH TO SEA-WATER MONITORING

Daniele D. Giusto, Luca Parodi, and Gianni Vernazza

Dept. of Biophysical and Electronic Engineering,
Univ. of Genoa, Via Opera Pia 11A, I-16145 Genoa, Italy

ABSTRACT

In this paper, we present some preliminary results obtained by using a system able to integrate data provided by remote sensing stations and oceanographic ones. The aim is to monitor the total suspended-matter content of sea-water (i.e., sea-water quality).

Attention is focused on some image-processing tools, the final purpose being the creation of a sea-water quality map based on Landsat data.

A segmentation technique using rank-order operators is presented: it allows a simultaneous analysis of data provided by different Landsat spectral bands.

Results are discussed, and the work currently in progress on the fusion of such results with those of an oceanographic analysis is described.

KEYWORDS: data fusion, image segmentation, sea-water quality.

INTRODUCTION

Due to its close proximity to urban and industrial areas, sea-water is subject to increasing pollution.

Remote-sensing techniques are powerful investigation tools for rapidly acquiring data and information about sea-water quality and coast dynamics.

Such techniques make it possible to carry out the following geological and environmental investigations:

- study of sea-water quality in relation to river discharges and to the influence of densely populated areas;
- dispersion of sediments from river mouths into sea-water;
- monitoring of coast urbanization;
- extension and erosion of shores as a consequence of coast dynamics (currents and wave motion);
- mapping of submerged vegetation.

In order to study coastal environments, some methodologies have been developed

that are able to utilize multiple-sensor images acquired by a satellite.

Such techniques constitute the main investigation tools, which should be calibrated and supported by "in situ" measurements.

In this paper, we present some preliminary results concerning a system for the integration of oceanographic data on the total suspended-matter content of the sea (as well as on the differentiation of such content between organic and mineral substances), into data provided by high-resolution satellite images (e.g., Spot and Landsat sensors).

The main objectives of this research are:

- to complete the information derived from "in situ" measurements on dispersion and accumulation of sediments, evaluating the surface distribution of these materials to a better approximation;
- to interpret the satellite images, on which maps of the total suspended-matter content of coastal waters are based.

Integration of "in situ" measurements into remote-sensing data seems to be the best procedure for monitoring the quality of coastal waters.

A first problem to be solved, when working on different kinds of images, is posed by the different resolutions for such images. Hence, we have to adjust data in order to reach the same resolution for each image. The major problem is one of obtaining a symbolic description (e.g., by the segmentation process) of a scene acquired by multiple sensors.

In other words, we must analyze the properties of each pixel in the different images, and try to merge them appropriately (by a data-fusion approach), if their values are within a predefined narrow range, thus obtaining regions with uniform characteristics.

Toward this end, we have developed a segmentation technique, that is based on rank-order operators [Zamperoni, 1988], and that groups neighbouring pixels with similar spectral properties, examining the local histograms of each image.

The output of this process (i.e., division of the observed zone into some regions) constitutes the basis for the construction of a sea-water map.

In the following, the segmentation algorithm is described, and some results are reported, that have been obtained by working on Landsat images of the Ligurian Sea.

Moreover, a comparison with oceanographic data is made, and the research work currently being done on the integration of all such data is outlined.

METHOD OF ANALYSIS

The first processing step is aimed at reducing the noise at high spatial frequency that affects Landsat images.

In the present application (i.e., analysis of sea-water), no fine details are present so that we can try to eliminate noise effects, without addressing the problems that usually arise from smoothing edges or thin structures.

Many smoothing techniques operate in such a way as to generate new grey levels: this characteristic must be regarded as a drawback.

In fact, for image segmentation purposes, no new grey-level values should be generated by the operators used.

The path followed to meet this requirement leads to use rank-order operators, which have proved very useful in image smoothing.

In particular, a well-known example of this kind of filters for applications to image processing is the median filter.

We focused our attention on the so-called "representative grey value" (RGV), which can be determined by analyzing the local grey-level histogram.

The algorithm works as follows:

- an $n \times n$ square processing window is considered for each pixel in the original image;
- a histogram is derived from the grey values of the pixels belonging to the window;
- the RGV is extracted from the histogram, and replaces the grey level of the central pixel in the window.

Roughly speaking, the RGV may be regarded as the grey level associated with the maximum of the histogram.

If there is no single maximum, the algorithm performs a local analysis in the neighbourhood of each maximum: the maximum whose mode contains more pixels than the others is taken as the RGV.

This procedure allows a sensible noise reduction, together with a notable edge enhancement.

At this point, the segmentation step is performed by a data-fusion approach.

In the present application, we employ a region-growing operator that analyzes a larger number of different data (i.e., images).

The basic idea of the region-growing operation is to make a region growth

around its starting points, applying a computed threshold to make comparisons between the pixel under investigation and the neighbouring ones.

These comparisons allow the pixel to be properly labelled.

In our opinion, the use of a single threshold for the whole image might be regarded as a drawback, which, in our system, can be overcome by means of locally computed thresholds.

Data coming from the different spectral bands are utilized in a parallel way in order to label pixels.

The whole approach may be viewed as a kind of low-level data fusion.

In more detail, the proposed system works as follows:

- each input image is divided into square processing windows of 32×32 pixels; this window dimension is a good compromise between a short running time (which requires a larger window dimension) and the need for carrying out an analysis that is as local as possible (which requires a smaller window dimension);
- for each processing window, a grey-level histogram is computed, and, for each histogram, the main peaks are singled out. As a consequence, a threshold is computed between each pair of adjacent peaks (in this sense, we can say that each pixel is located between its two thresholds); this peculiarity is typical for each input image. The use of a different pair of thresholds for each pixel is a novelty, and represents an adaptive characteristic of the algorithm;
- each pixel is labelled according to its thresholds to allow comparisons between the pixel being labelled and its neighbouring pixels that have already been labelled. Such comparisons may give rise to some different situations:
 - the pixel may be regarded as the starting point of a new region (new label);
 - the pixel belongs to an adjacent region (propagation of an existing label);
 - the pixel acts as a "bridge" between two different regions (merging of these two regions and inclusion of the current pixel in the resulting region).

The outcome of the segmentation process is a large number of small regions; therefore, a merging step is required;

- to perform the merging step, each located small region is examined, and may be merged into a larger one if some similarity criteria are complied with. Subsequently, more restrictive similarity criteria are considered and applied to merge regions with quite large areas.

All these criteria are derived from the information available, without weighting different parameters in different ways.

This allows small regions to be preserved, even when the significance of a small region is based on only one parameter.

RESULTS

The described algorithms have been used to analyze a Landsat MSS image of the Ligurian Sea (Fig. 1 presents the global scene). In the following, we consider a zone around the estuary of the Arno river. To analyze sea-water quality, we have utilized bands 1 and 2: because of the low resolution in the grey levels of the sea-water (about 20 levels), we have changed the grey scale to facilitate visual analysis by a human operator.

Figs. 2 (band 1) and 3 (band 2) show the zone examined.

Region-growing segmentation has been performed after the filtering step, working simultaneously on the two images: it has furnished the map displayed in Fig. 4.

The present approach requires a very short running time: globally, it takes about one minute of CPU time on a Hewlett Packard 835 TurboSRX workstation.

CONCLUSIONS

The proposed data-fusion approach to image segmentation has proved a powerful tool for low-level analysis of remotely sensed images.

The system presented in this paper is the first step towards the integration of oceanographic data into remote-sensing images, the aim being the monitoring of sea-water quality.

At present, we are working on such information fusion, utilizing both the results obtained by the segmentation process and data coming from oceanographic campaigns (Fig. 5 gives an example of this kind of data for the same zone as in Figs. 3 and 4).

Future developments will concern the analysis and integration of pictures taken in different seasons, or provided by different sensors (e.g., Spot, SAR, etc.).

ACKNOWLEDGMENT

The authors thank Dr. A. Ramella (Dept. of Earth Sciences, University of Genoa, Italy) for supplying Landsat images and oceanographic data.

REFERENCE

1. Zamperoni, P., "Feature extraction by rank-order filtering for image segmentation", *Int. J. of Pattern Recognition and Artificial Intelligence*, Vol. 2, No. 2, pp. 301-319, 1988.



Fig. 1. The global Landsat MSS image of the Ligurian Sea used in testing the system. An expansion of the grey scale has been performed to facilitate visual analysis by a human operator.



Fig. 2. A portion of the image in Fig. 1: it represents the estuary of the Arno river (band 1).

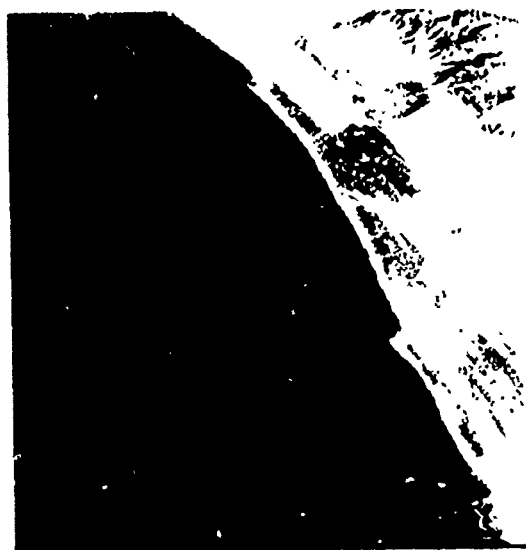
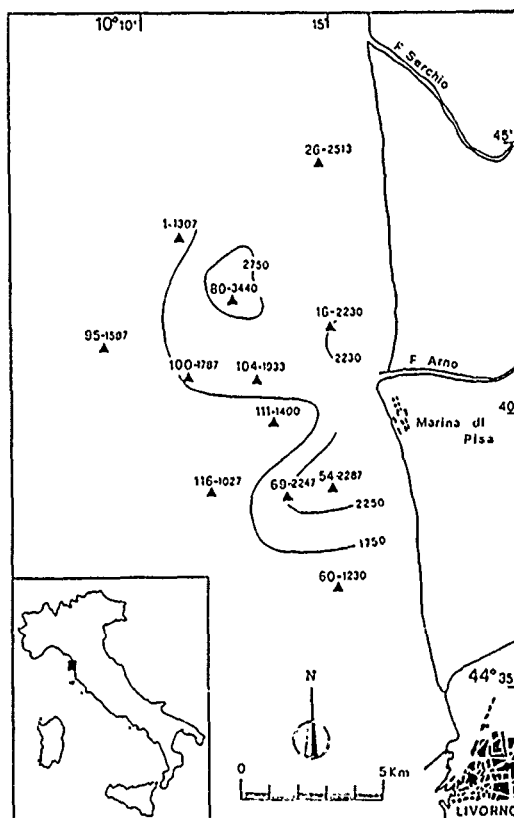


Fig. 3. The same zone as in Fig. 2 (band 2).



Fig. 4. The result of the segmentation process performed by fusing the two information channels.



| n Camp | Lat. | Long. | %T/m | T.P. | M.P. | O.P. | % M.P. | %O.P. |
|---------|----------|----------|------|------|------|------|--------|-------|
| F. ARNO | | | | | | | | |
| 1 | 43°43',2 | 10°11',0 | 32 | 1307 | 650 | 657 | 49,7 | 50,3 |
| 16 | 43°41',6 | 10°15',0 | 32 | 2230 | 1325 | 905 | 59,4 | 40,6 |
| 26 | 43°44',6 | 10°14',6 | 34 | 2513 | 1186 | 1327 | 47,2 | 52,8 |
| 54 | 43°38',7 | 10°15',1 | 25 | 2287 | 1217 | 1070 | 53,2 | 46,8 |
| 60 | 43°36',8 | 10°15',3 | 58 | 1230 | 623 | 607 | 50,7 | 49,3 |
| 69 | 43°38',5 | 10°13',8 | 26 | 2247 | 1200 | 1047 | 53,4 | 46,6 |
| 80 | 43°42',1 | 10°12',3 | 20 | 3440 | 1567 | 1873 | 45,5 | 54,5 |
| 95 | 43°41',2 | 10°09',0 | 31 | 1587 | 807 | 780 | 50,8 | 49,2 |
| 100 | 43°40',7 | 10°11',1 | 27 | 1787 | 900 | 887 | 50,4 | 49,6 |
| 104 | 43°40',6 | 10°13',1 | 24 | 1933 | 866 | 1067 | 44,8 | 55,2 |
| 111 | 43°39',8 | 10°13',3 | 33 | 1400 | 840 | 560 | 60,0 | 40,0 |
| 116 | 43°38',6 | 10°11',8 | 35 | 1027 | 467 | 560 | 45,5 | 54,5 |

Fig. 5. An example of oceanographic data: they refer to the same zone as in Figs. 2 and 3.

UTILITY OF AIRBORNE ELECTROMAGNETIC PROFILER DATA FOR DETERMINING WATER DEPTHS AND CHARACTERIZING COASTAL SEDIMENTS AND MARSH SOILS

R.E. PELLETIER AND S.T. WU

NASA-SCIENCE AND TECHNOLOGY LABORATORY
JOHN C. STENNIS SPACE CENTER
STENNIS SPACE CENTER, MS 39529

ABSTRACT

Many of today's typical airborne or spaceborne remote sensing devices are of limited use in studying soils and wetland sediments because most of these devices are relatively surface-oriented. In wetland environments, not only is the subsurface inaccessible, but the sediment surface is also obscured by a layer of water. The Naval Ocean Research and Development Activity recently developed a helicopter airborne electromagnetic (AEM) system for mapping bathymetry in shallow marine environments, but this system also has utility in assessing a number of physical properties of both the water and underlying sediment. The system contains an electro-magnetic source for the production of multi-frequency eddy currents in the underlying water and sediment material, a receiver for detection of secondary electromagnetic fields produced by the primary currents, and a recording system to measure the in-phase and quadrature levels of the secondary currents. Inversion techniques are used to determine a variety of variables including depth of water, water and sediment conductivities, and sediment density and porosity. This study gathered and analyzed AEM data for a region around Cape Lookout, North Carolina. Preliminary results illustrate great promise for the AEM as a tool to map bathymetry and characterize at least some physical properties of water and submerged materials.

(Key words: electromagnetic method, bathymetry, salinity)

INTRODUCTION

The method as discussed in this study, often referred to as the induction electromagnetic (EM) method in mining geophysics, involves the propagation of a time-varying low-frequency EM field in and over the earth. This technology in its early forms started in the beginning part of this century and has been applied to a variety of geological problems.

Typically, and especially early on, only single frequencies were used. Later, wide-band multi-frequency sources were employed (Ryn et al., 1972, and Ward et al., 1977). Drawbacks of

such systems employing multiple frequencies is that each frequency requires separate operation, making it difficult to acquire simultaneous, multi-frequency data from a moving platform. Therefore, initial efforts to make such a system airborne were through the use of only one frequency (Fraser, 1978). Multi-frequency systems soon followed, such as the one used by Won and Smits (1986) for early studies in bathymetric charting of shallow marine environments.

In response to these findings by Won and Smits, the Navy built a multi-frequency (three frequencies tunable between 90 and 4050 Hz) system with fast frequency - multiplexing and power-switching to enable data acquisition from a moving helicopter platform. While this system was primarily developed for depth charting purposes, much can be learned about subsurface soil and sediment properties in very shallow, inundated environments such as marshes, swamps and beat bogs. Further analysis indicated the ability to penetrate the submerged sediments and determine some sediment physical properties as well.

This study acquired AEM data over a coastal region of North Carolina near Cape Lookout with the recently developed NORDA AEM system as part of a collaborative prototype testing program. This paper will present some of the preliminary results from that testing program as a means to begin an evaluation of the utility of the sensor's ability to measure water depths in shallow marine environments as well as to characterize a number of water and submerged sediment physical properties.

APPROACH

Site Description-The study site for this investigation is located South of the Pamlico Sound near the city of Beaufort, North Carolina and extending seaward around Cape Lookout. The area is dominated by two barrier island chains, coming together at a near right angle to form Cape Lookout, a migrating recurved spit enclosing a quiet water lagoon. This lagoon serves as a sediment trap for fine-grained clays and organic rich, suspended particles exiting to the ocean from the Back Sound. Sediment types vary from predominantly organic

materials and clays near the marshy regions to sand in the Back Sound to sands and gravelly materials on the ocean side.

The Back Sound is a relatively shallow area behind the barrier island averaging about three feet in depth at mean low water, but ranging from emergent *Spartina alterniflora* marshes to natural and man-made channels up to about 25 feet in depth. The ocean side of the barrier island drops to 40-50 feet in depth within a mile. Data for this study were collected up to 60 feet depth several miles off the coast. Salinity of the water ranges from brackish near the small upland drainage rivers to saline in the Back Sound and in the ocean.

System Specifications and Data Acquisition- System electronics are composed of sensor electronics (e.g., transmitter, receiver, bucking coils) and onboard aircraft electronics (e.g., signal processor, graphic printer, display, recorder, and calibration). For this study we employed the 270, 870 and 4050 Hz frequencies since we wanted to focus on the more shallow regions of study site and the higher frequencies provide for more shallow field penetration. Data was acquired in-phase and out-of-phase at the three frequencies.

The helicopter system was flown at a speed averaging 70 knots at a height of around 250 feet based on altimeter readings. A towline kept the sensor suspended about 100 feet below the center of the aircraft averaging a distance of approximately 150 feet between the water or land surface and the sensor. The data rate was set at 30 readings per second and preprocessed to produce an average reading per second. The resulting data produces a resolution size of 30-40 m across track (dependent upon frequency and sensor height). Along track, the resolution can be considerably better (dependent upon data acquisition speed and sampling rate). The system was calibrated for baseline drift at a height of 1500 feet. Maximum drift rates typically did not exceed 5 ppm per hour.

Flight Line Design- More than 300 km of flight lines were flown in a series of equi-spaced parallel transects approximately 250 m apart. An almost near North-South orientation was chosen for the flight lines to minimize Earth's natural magnetic forces within a line.

The lines were designed to cross gradients in salinity, topography/bathymetry, and sediment type (e.g., organic vs. mineral). Two focal areas were chosen for illustration in this paper. The first is an approximately 16 km line flown from a shallow brackish region in the North River southward across channels, marshes, the barrier island and into the open ocean. The other focal area encompasses the Lookout Bight which demonstrates quick and highly variable changes in bathymetry and sediment type.

A mini-ranger system was employed for geopositioning the data. Three transmitting stations were strategically located across the test site and the receiver was placed onboard the helicopter. Geographic coordinates were recorded and merged with the AEM data to

provide locational information for the measured data on a per second basis.

Data Processing- Data inversion techniques similar to the classic Newton-Raphson method were employed. In this initial evaluation of data from the prototype AEM system, the following five variables were derived: 1) water depth; 2) water conductivity; 3) sediment conductivity; 4) sediment density; and 5) sediment porosity.

RESULTS

Figure 1 illustrates data from the approximately 16 km long North-South line (right to left in the graph) extending from the North River out into open ocean. Notice the shallow bathymetry in the northern section of the line and the channels as one approaches the barrier island. "0" readings for depth indicate emergent land. Bathymetry drops quickly on the ocean side and illustrates many small shoals. Water conductivity and sediment conductivity exhibit wide ranges in variation in the more shallow regions of the Back Sound and marshy areas than on the ocean side. The heterogeneous sediment types and shallow depths cause much of this variation. Water in shallow tidal pools were measured up to 5°C higher in temperature and 2-3 times higher in conductivity than water in more open areas because of the impact from the heat of the sun and evaporation in these places. The open ocean water is more homogeneous in composition as is the ocean's sediment composition.

Figure 2 illustrates water depth measurements for four consecutive transects cutting across the Cape Lookout Bight. 2a illustrates each transect independently. The "X's" indicate ground truth depth measurements from 1:40K scale bathymetric maps. Correlation appears exceedingly good between the AEM measurements and ground truth except for a small segment within one of the cavernous features along line 122. We feel that the AEM measurements are probably a more accurate depiction of the actual depth at the time of acquisition since the map indicates that this is a very unstable region of rapidly changing shoals. 2b provides a somewhat better 3-dimensional perspective of the region.

REFERENCES

1. Fraser, D.C., "Resistivity Mapping with an Airborne Multicoil Electromagnetic System," *Geophysics*, Volume 43, Number 1, pp. 144-172, 1978.
2. Ryn, J., H.F. Morrison and S.H. Ward, "Electromagnetic Depth Sounding Experiment Across Santa Clara Valley," *Geophysics*, Volume 43, pp. 351-374, 1972.
3. Ward, S.H., D.F. Pridmore and L. Rijo, *NSR Workshop in Mining Geophysics*, University of Utah, p. 309, 1977.
4. Won, I.J. and K. Smits, "Characterization of Shallow Ocean Sediments Using the Airborne Electromagnetic Method," *IEEE Journal of Oceanic Engineering*, Volume OE-11, Number 1, pp. 113-122, 1986.

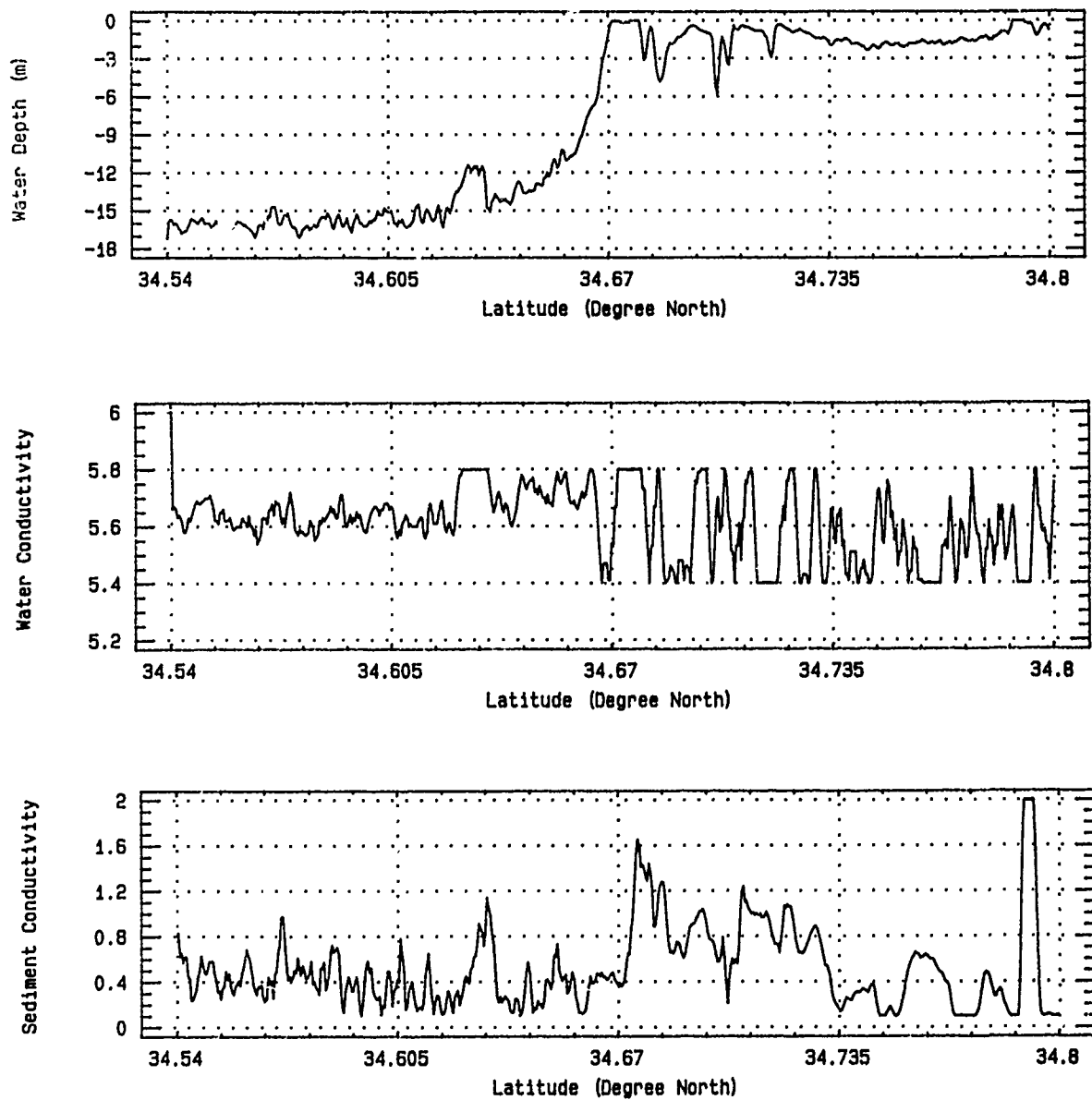
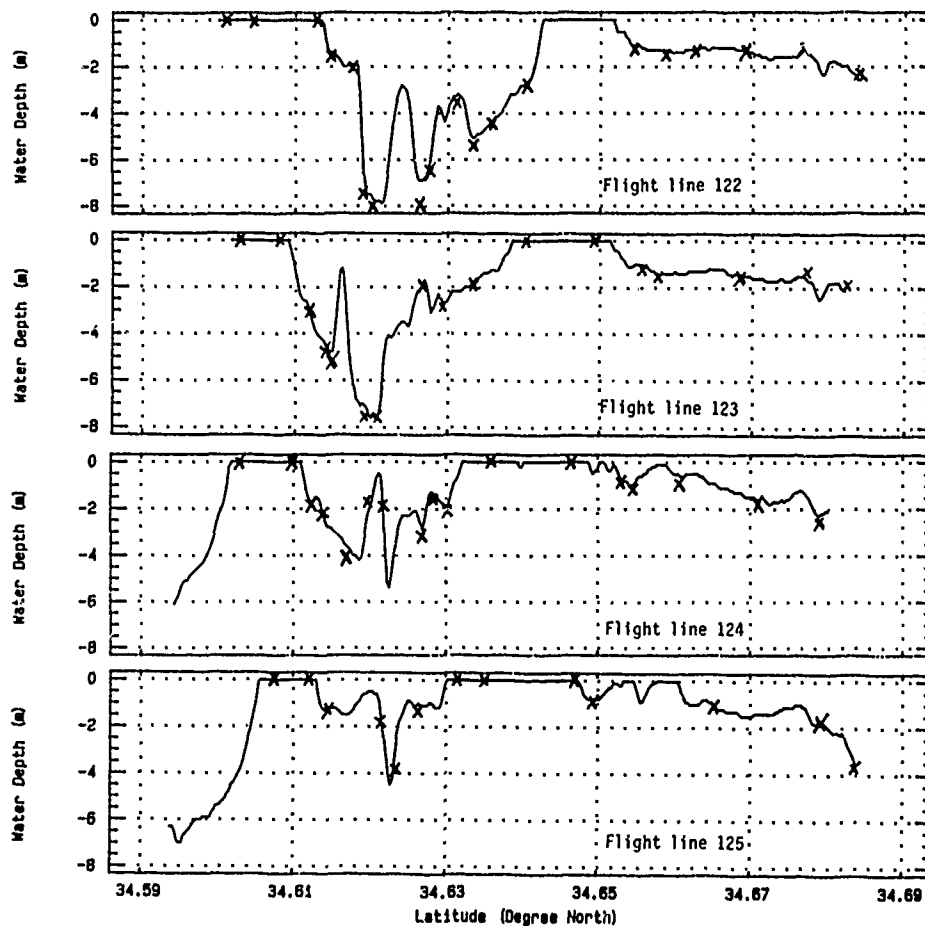


Figure 1. Three inversion technique - derived variables - depth of water at acquisition, water conductivity, and sediment conductivity - for line 301 depicted along a latitudinal gradient on the X-axis.

2a)



2b)

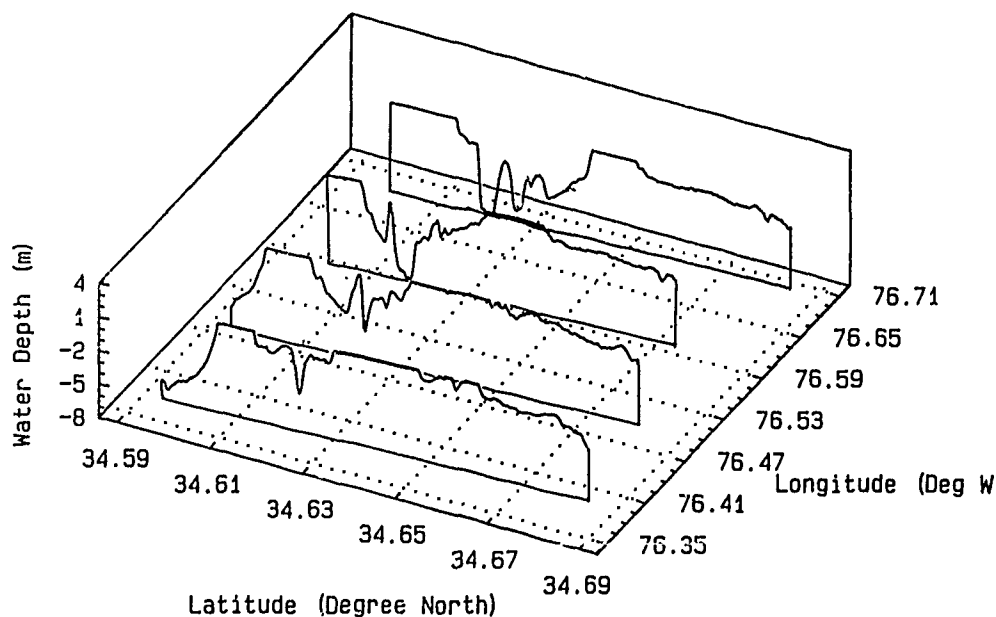


Figure 2. Inversion technique - derived depth of water at acquisition variable for four consecutive transects through Cape Lookout Bight. "X's" indicate ground truth depth values as determined by NOAA 1:40K bathymetric maps in 2a. 2b provides a three-dimensional perspective of these same four transects.

Bathymetry retrieval algorithm for coastal water [case II]
from Landsat - 5 Thematic Mapper Data

A.K. Mishra, P.N. Sridhar and K.S. Prasad

National Remote Sensing Agency
Balanagar, Hyderabad - 500 037
INDIA

Abstract:

Bathymetry retrieval by satellite remote sensing in case II waters was found inefficient using the existing relationships between attenuation coefficient of water, bottom reflectance, total incident radiation and upwelling radiance. A study in the case II waters off Paradip, Bay of Bengal, where average diffused attenuation coefficient $k(0.55\mu\text{m}) = 0.45$ per meter) showed a linear relationship between the reflectance (R) and the bottom depth (D). In the present work a satellite synchronous sea truth study was made to develop a relationship between reflectance (R) and depth (D) to predict near shore bathymetry using a statistical correlation. In this procedure we deviate, from the routinely used deep water signal subtraction method for atmospheric correction, by augmenting it using the clear water radiance approximation. This algorithm can be applied to coastal waters where the attenuation coefficient is around 0.5 per meter at $0.55\mu\text{m}$.

Some byproducts of this study are : (a) an indication of the probable loss in the sensitivity of TM bands (b) suitability of band II ($0.57\mu\text{m}$) in presence of diverse atmospheric constituents. In the present study we have attempted to use a general value for Angstrom exponent for determination of aerosol optical thickness which can be applied over both marine and coastal environments. The results have been very encouraging.

KeyWords: Reflectance, bottom depth, Angstrom exponent, sensitivity, TM calibration.

Introduction:

Bathymetry charting in particular is very time consuming and expensive as there is a need for soundings to properly delineate the depth variations due to rocks, shoals and sediment which are potential navigational hazards. With the advent of satellites intense interest have been shown in remote sensing methods to

compliment or replace conventional methods and monitoring changes in nearshore topography. Essentially the passive remote sensing technique depends upon finding a relationship between the measured water depth and reflected radiance.

Hallada (1984) has shown that Landsat 4 and 5 have good capability for mapping bathymetry and bottom cover, particularly in clear ocean areas. The maximum detectable water depth with his algorithm was 21m in the blue band. He concludes that variations in water attenuation properties dominated residual errors in predicted water depths less than 8m. The success of remote bathymetry further lies in the water extinction coefficient. In clear water (extinction coefficient = 0.058 m^{-1}) depths upto 22m could be measured reliably and signals from 40m depths could be differentiated sometimes from deep water signals that are due to scattering only (Polyn, 1976).

Most of the earlier workers have proposed bathymetry retrieval algorithms for clear waters (case I), but majority of coastal waters fall in case II where suspended sediments and dissolved organic matter dominate and interfere in the upwelling signal. There was an immediate need for bathymetry retrieval algorithm development in coastal waters (case II). We have attempted to study the reflectance of coastal waters of Paradip, India and present an algorithm for bathymetry retrieval. Sea-truth collections were performed synchronous to Landsat (139-046) pass over Paradip. Depth and water quality parameters were collected using a pilot boat of the port authorities. Water samples were analysed for chlorophylla and suspended sediment concentrations. A relationship was developed using fifty-nine depth points out of which nine were collected synchronous to the satellite pass and the rest were extracted from a recent hydrographic chart and corresponding atmospherically corrected reflectance values from Landsat TM Band 2. We have deviated from the usual practise of employing the deep water signal subtraction in the atmospheric correction procedure as substantial information on

upwelling irradiance was lost when dealing with coastal waters. Atmospheric correction indicates that there may be a loss in the sensitivity of the thematic mapper detector (see Table II). Our studies reveal that single Band (TM 2) can be reliable for coastal bathymetry mapping provided accurate corrections for atmospheric influences are performed on the data. The algorithm needs further validation in waters of different attenuation coefficients.

2. Atmospheric correction:

Atmospheric corrections for TM were based on CZCS procedures (Gordon, 1983). The radiance at sensor $L(\lambda)$ is given by

$$L(\lambda) = L_R(\lambda) + L_A(\lambda) + L_W(\lambda) * TRA_{OZ}(\lambda, \mu) \quad \dots 1$$

Where $L_W(\lambda)$, $L_R(\lambda)$ and $L_A(\lambda)$ are water leaving radiance, radiance due to Rayleigh and aerosol scattering respectively. $TRA_{OZ}(\lambda, \mu)$ is the irradiance transmittance through the atmosphere $\mu = \cos$ of view zenith

$$TRA_{OZ}(\lambda, \mu) = \exp[-\frac{1}{2}tr(\lambda) + \beta_A * ta(\lambda) + toz(\lambda)/\mu] \quad \dots 2$$

where tr , ta , toz are the optical thicknesses of Rayleigh, aerosol and ozone respectively, and β_A is the aerosol backscattering coefficient (Tanre, 1979). For completeness of procedure the determination of computed terms in equation 1 are given below (Gordon, 1983)

$$L_R(\lambda) = \frac{W_R(\lambda) \cdot P_R(\lambda, \psi) \cdot E'_o(\lambda)}{4\pi} \quad \dots 3$$

where $W_R(\lambda)$, $P_R(\lambda, \psi)$ and $W_A(\lambda)$, $P_A(\lambda, \psi)$ are the single scattering albedo and scattering phase function for Rayleigh and aerosol respectively.

$$P_R(\lambda, \psi_-, \psi_+, \lambda) = [P_R(\lambda, \psi_-, \lambda) + (p(\theta) + p(\theta_o))] * P_R(\lambda, \psi_+, \lambda) / \cos \theta \quad \dots 4$$

where the solar radiation and the reflected radiation from the sea surface are scattered in the atmosphere at an angle ψ_- , ψ_+ respectively.

$$\cos \psi_- = -\cos \theta * \cos \theta_o - \sin \theta * \sin \theta_o * \cos(\phi - \phi_o) \quad \dots 5$$

$$\cos \psi_+ = \cos \psi_- + 2 \cos \theta * \cos \theta_o \quad \dots 6$$

$$\text{and } t_R(\lambda) = 0.00879(\lambda)^{-4.09} \quad \dots 7$$

$$E'_o = E_o \exp[-toz(1/\cos \theta_o + 1/\cos \theta)] \quad \dots 8$$

$P_R(\psi) = (3/4) * [1 + \cos^2 \psi] \quad \dots 9$
The angles needed for atmospheric correction were computed from the following parameters (Sturm, 1983)
 θ = zenith angle of satellite at pixel location

ϕ = satellite azimuth in pixel location
 θ_o = sun zenith angle at pixel location

ϕ_o = sun azimuth at pixel location

2.1 Angstrom component computation:

For atmospheric correction procedures the water leaving radiance over clear oceanic waters is considered ≈ 0 (Band 3 ; central wavelength in TM is 660 nm). W_R and W_A were taken as unity. Aerosol particles were considered as non-absorbing spheres.

Sturm (1980) discusses methods of computing which requires a knowledge of the local meteorological visibility as a measure of the amount of aerosol haze present. However some research indicates that the wavelength dependence of ta is given by:

$$ta(\lambda) = \beta(\lambda) * \alpha \quad \dots 9$$

In the present study an attempt has been made to compute the value of Angstrom's slope α and to present an value for general applications. This was extrapolated from studies at Hanford Meteorological station (HMS) and Rattle snake mountain observatory (Michalsky, 1982) for 13 days in September to October. A mean value of $\alpha = 1.375$ was thus computed. This α was used for computing the optical thickness for two CZCS scenes. The comparison of computed and observed Angstrom values are shown in Table 1. The residual error obtained was negligible. We further conclude that the $\alpha = 1.375$ can be used for computations of aerosol optical thickness in coastal zones, where continental and maritime air masses blend.

2.2 Retrieval of reflectance $R(\lambda)$:

The scattering phase function for aerosol $P_A(\psi, \lambda)$ has been computed from normalized scattering function for aerosol (Quenzel, 1983) considering $v = 3$ for aerosol type 'marine surface'. Using equations 1, 2 and 3 the aerosol optical thickness at 0.660 μm has been determined over clear ocean water. The optical thickness thus obtained was used for determination of aerosol optical thickness at 0.485 μm (central wavelength of TM Band 1) and 0.570 μm (central wavelength of TM Band 2). The computed radiance value and the radiance values at the sensor have been presented in Table 2 for comparison. It can be seen that there is a fall in the radiance values of Band 1 and 2 of TM for clear water. This may be due to loss in sensitivity of the detector of Landsat TM.

The radiance at the sensor $L_r(\lambda)$ has been computed using

$$L_r(\lambda) = [(N(\lambda) - b(\lambda)) / g(\lambda)] * C(\lambda) \quad \dots 10$$

Where $N(\lambda)$, $g(\lambda)$ and $b(\lambda)$ are the digital counts, gain and bias respectively and $C(\lambda)$ is the calibration correction determined from the approximation

of clear water radiance. Table 3. gives the computed values along with the mean extra terrestrial solar irradiance computed by weighing the Neckel and Labs (1981) $E_0(\lambda)$ with the spectral response of TM.

The reflectance $R(\lambda)$'s were readily computed from the retrieved $LW(\lambda)$'s as

$$R(\lambda) = \frac{\pi n^2 LW(\lambda) \dots 11}{(1 - \rho) \mu_0 E_0 \text{TRAQ}(\lambda, \mu_0)}$$

$n = 1.341$ and $\rho = 0.021$.

Where n is the refractive index of water and ρ is the Fresnel's reflectivity. For validation of the atmospheric correction thus developed a chlorophyll retrieval algorithm [chl_a conc. mg/m³ = 2.76PI - 0.45 where PI is the ratio of TM band I and Band II] by the authors (in prep) has been applied and the TM estimated and ship measured concentrations have been presented in Table 4. It can be seen that the atmospheric correction thus developed has worked efficiently.

3. Image Analysis:

Image analysis was performed on improved multispectral data analysis system - DIPIX which runs on ARIES userfriendly software. Cloud free thematic mapper data (Path 139, Row 046) of 16 January, 1988 was used for the analysis. As sea - truth data viz; depth, suspended sediments and chlorophyll concentrations were collected synchronous to the satellite pass - our aim was to locate the station positions on our image using a base map. The base map with station locations and depths was filmed and later digitized using a drum scanner. The digitised CCT was later loaded on to the image processing system for further analysis. We used the base map as the master and our image as a slave to rotate the image. To facilitate this task we located ground control points on the slave and master images. A polynomial transform (PT) file was created along with residual error listings. This PT file was used in a resampling task to rotate the image. This task took care of the geometric correction.

3.1 Radiometric Correction:

Radiometric corrections were performed on VAX 11/780 system. The radiance contribution due to all atmospheric constituents were computed and the appropriate corrections for band 1 and band 2 were performed. Band ratioing was performed for applying the chlorophyll retrieval algorithm and validating the atmospheric correction.

The radiometrically corrected band 2 data was used for locating the stations. The corresponding radiances were computed and a regression was performed with the known depths. (See Fig.1). The relationships showed a negative correlation. Using the equation

a look up table (LUT) was computed for pseudo color density slicing for showing variations in depths. Fig. 2

Results and Discussion:

An exotech TM field radiometer was used to measure the downwelling irradiance during satellite pass time. The results are tabulated in Table V. A comparison between the calculated and observed downwelling at sea surface shows considerable variation. The overall estimation of transmittance do not agree with the observed values particularly at 0.485 μ m and 0.660 μ m than at 0.570 μ m. This indicates that the band II information is less noisy and more reliable for bathymetric studies when compared to band I and III. The results support the view that single band (TM band 2) information can be used for coastal bathymetric mapping. Though earlier workers have concluded that hybrid multiple TM band regression algorithms are superior to exclusive single band regressions, our results with coastal waters of Paradip have been encouraging and have proved that it could be used for such applications. The usual practice of employing the deep water signal subtraction in the atmospheric correction procedure would suppress the upwelling radiance in coastal waters. This is true when we relate the total reflectance (R) to the depth. The computed L_r and L_a values have been subtracted from the total L_r values all over the scene instead of deep water signals, considering a homogenous aerosol distribution.

The extension of this algorithm to different coastal waters requires further testing and validation. This is due to the variations in attenuation coefficients encountered at different locations. A byproduct of our atmospheric correction was information on the probable loss in the sensitivity of the TM sensor. We have used the clear water approximation method to arrive at such conclusions. We conclude that single band (TM band 2) has a good capability for mapping near shore bathymetry particularly in coastal (case 2) waters where there is an acute need for such bathymetric profiles.

References:

1. Hallada, W.A. - Bathymetry analysis using Landsat-4 Thematic Mapper data. Science applications research, MD, 35 pp. (1984)
2. Polyn, F.C. - Final report on NASA/cousteau Ocean Bathymetry Experiment - Remote Bathymetry using High Gain Landsat Data, NASA-CR-ERIM-118600-1-F, 131 pp (1976).
3. Sturm B. Selected topics of Coastal zone color scanner (CZCS) data evaluation. In: Remote sensing applications in marine science and technology Ed: A.P. Cracknell, D. Reidel publishing company, Dordrecht, 137-168 (1983).

4. Tanre D., Herman M., Deschamps P.Y., DeLeffe A., Atmospheric modelling for space measurements of ground reflectances including bi-directional properties. App. Optics. 18 21pp 3587 - 3594 (1979).
5. Howard R. Gordon., Dennis K. Clark., James W. Brown., Otis B. Brown., Robert H. Evans and William W. Broenkow Phytoplankton pigments concentrations in the Middle Atlantic Bight : Comparisons of ship determinations and CZCS estimates. App. Optics. Vol 22 No 1 p 20-36 (1983).
6. Sturm B. The atmospheric correction of remotely sensed data and the quantitative determination of suspended matter in marine water surface layers, In: Remote sensing in Meteorology, Oceanography and Hydrology Ed. A.P. Cracknell Ellis Horwood Ltd (1981).
7. Quenzel H. scattering , Absorption , Emission and radiative transfer in the atmosphere In: Optical remote sensing of Air pollution Elsevier science publishers 25pp (1983).
8. Michalsky J.J., Laulainen N.S., Stokes G.N., and Kleckner E.W Atmospheric aerosols, their formation, optical properties and effects Ed: Adarsh Deepak, Spectrum press (1982).

Fig. 1 Reflectance (R) vs Depth (D)

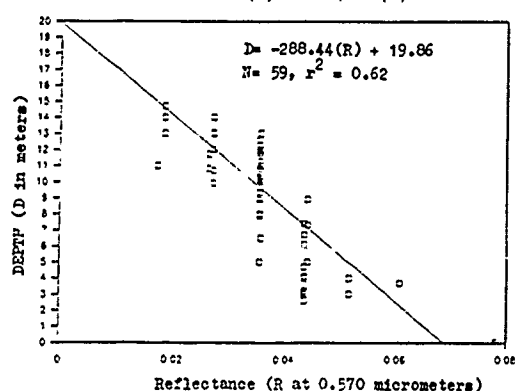


Fig. 2 DIPIX output of classified image showing depth variations.

| Wave Length (µm) | t ₁ ^a | t ₂ ^a | residual error | t ₁ ^{**} | t ₂ ^a | residual error |
|------------------|-----------------------------|-----------------------------|----------------|------------------------------|-----------------------------|----------------|
| 0.443 | 0.181 | 0.177 | -0.022 | 0.364 | 0.406 | 0.115 |
| 0.520 | 0.157 | 0.142 | -0.096 | 0.317 | 0.326 | 0.028 |
| 0.550 | 0.124 | 0.131 | 0.057 | 0.277 | 0.301 | 0.687 |
| 0.670 | 0.100 | 0.100 | 0.000 | 0.230 | 0.230 | 0.000 |

Table I: ^a Computed aerosol optical thickness over Western British Channel, Orbit No. 17997, May 18, 1982; ^{**} Computed aerosol optical thickness over Senegal (Dakar), Orbit No. 16533, Date Feb 01, 1982; [†] Calculated aerosol optical thickness with $\alpha = 1.375$.

| Wave Length (µm) | L _N | L _A | T _{RAOZ} L _N | L _T Calculated | L _T from Sensor | Residual error |
|------------------|----------------|----------------|----------------------------------|---------------------------|----------------------------|----------------|
| 0.485 | 3.038 | 1.4270 | 0.4597 | 4.9247 | 3.8906 | -0.2658 |
| 0.570 | 1.434 | 1.0407 | 0.0559 | 2.5306 | 2.3450 | -0.0789 |
| 0.660 | 0.680 | 0.7460 | 0.0053 | 1.4313 | 1.4260 | -0.0037 |

Table II: L_N, L_A, L_N, L_T and L_T are in mW / cm² µm sr.

| Wave Length (µm) | g | b | C | E ₀ |
|------------------|-------|-------|-------|----------------|
| 0.485 | 15.55 | 1.830 | 1.266 | 194.40 |
| 0.570 | 7.85 | 1.690 | 1.079 | 183.40 |
| 0.660 | 10.20 | 1.880 | 1.000 | 154.60 |

Table III: g is in counts/(mW/cm²µm sr); b - counts; E₀ in mW / cm²µm.

| Station No. | Station Depth D | Chlorophyll Conc. Ship measurement C [*] | Chlorophyll Conc. From TM C ^{**} | Residual error | Att. Coeff (K) at 0.55µm | SSC observed |
|-------------|-----------------|---|---|----------------|--------------------------|--------------|
| 6 | 20.0 | 2.3 | 2.1 | -0.087 | - | - |
| 7 | 14.7 | 1.2 | 1.4 | 0.166 | 0.20 | 0.60 |
| 8 | 10.7 | 1.1 | 1.0 | 0.636 | 0.60 | 1.00 |
| 9 | 8.9 | 1.4 | 2.1 | 0.500 | 0.50 | 0.50 |
| 10 | 6.5 | 2.2 | 2.1 | -0.045 | 0.50 | 0.70 |

Table IV: D in meters, C^{*} and C^{**} are in µg/m³ and K in m⁻¹ SSC is the suspended sediment concentration in µg/l.

| Wave Length (µm) | E ₀ T _{RAOZ} (µW) Calculated | E ₀ Observed | Residual Error |
|------------------|--|-------------------------|----------------|
| 0.485 | 101.33 | 81.46 | -0.244 |
| 0.570 | 100.80 | 104.39 | 0.034 |
| 0.660 | 89.66 | 100.29 | 0.106 |

Table V: E₀, E₀ are in mW / cm² µm. E₀ is the downwelling solar irradiance at sea surface synchronous to satellite pass time measured by Exotech field Radiometer.

NORMALIZATION OF TIDAL FLUCTUATIONS FOR USE IN SATELLITE REMOTE BATHYMETRY

Wei Ji, Daniel L. Civco and William C. Kennard

Laboratory for Remote Sensing
Dept. of Natural Resources Management & Engineering
The University of Connecticut
Storrs, CT 06269-4087 U.S.A.

Abstract: Modeling in satellite remote bathymetry requires a set of registered calibration water depths to estimate parameters of the prediction model using regression analysis. However, obtaining adequate water depth data concurrent with the satellite overpass unavoidably results in errors for water depth prediction introduced by the temporal inconsistencies between a point-in-time image data set and a set of sample depths taken over a span time during which the tides fluctuate nearly continuously. To alleviate this problem, a time-normalization technique was developed using NOS tide tables to construct a full tidal curve using the one-quarter, one-tenth rule. This technique can significantly reduce the effect of tidal fluctuations on sample water depths so that the accuracy of water depth models can be improved.

Key words: remote bathymetry, satellite data, tidal effects, data normalization.

Introduction

Reducing the quantitative error in modeling has been a basic task for scientists involved in satellite remote bathymetry since the first studies were carried out in the early 1970's. Those studies were concerned mainly with spectral properties of bottom conditions and the structure of depth-to-bottom models as an aid in developing mathematical models for predicting water depth which are insensitive to variations of bottom conditions (Polcyn and Sattinger, 1970; Lyzenga, 1977, 1979 and 1981; Paredes and Spero, 1983; Spitzer and Dirks, 1986 and 1987; Clark *et al.*, 1987). In spite of the fact that tidal height variations introduce a compounding factor, it appears that little attention has been directed toward understanding the effect of such fluctuations on the modeling of satellite remote bathymetry. The objectives of our study were to investigate how tidal fluctuations affect the modeling process in satellite bathymetry and to develop an approach to compensate for such variations.

Tidal-affected modeling in satellite bathymetry

Tides are the daily or twice daily rhythmic rise and fall of sea level, or waves with a length of approximately half the circumference of the earth (about 20,000 km) (Ross, 1977). Semidiurnal tides, which occur most commonly, typically have a period of about 12 hours and 25 minutes. Tides along the coasts of the world generally range between 1 and 3 meters, but can be much higher than that in some coastal areas such as the Bay of Fundy in Canada and the Bay of Hangzhou in China due to the geographic position and geometry of the area.

Satellite remote bathymetric techniques can be applied to coastal shallow water areas, most of which are significantly affected by daily tides. Operational models of satellite bathymetry

commonly used for water depth predictions in such areas are essentially statistical and are often based on radiance data from two or three discrete portions of the spectrum (commonly referred to as two and three-band models). Regardless of any theoretical considerations on which they are based, they will finally appear in a regression form such as the following two and three-band models (Clark, *et al.*, 1987; Spitzer and Dirks, 1986):

$$Z = A_0 + A_1 \ln X_1 + \frac{1}{2} \ln X_2$$

and

$$Z = A_0 + A_1 \ln X$$

where Z is the predicted water depth; X_1 and X_2 in the first model are combinations of the image brightness of the pixel of interest and the deep water in two different spectral bands; X in the second model represents such a combination, but of three different bands; A 's, the parameters of the model, are the combination of certain optical parameters and unpredictable factors. Usually, a set of registered calibration water depths is required in regression analysis to estimate parameters of the model. However, the process of obtaining an adequate sample of water depth data concurrent with the satellite overpass unavoidably results in errors for water depth prediction introduced by the temporal inconsistencies between a point in time image data set and a set of sample depths taken over a span of time during which the tides fluctuate nearly continuously. In practice, sample water depths are recorded over a several-hour period and, in fact, perhaps over several days, and almost all measurements will not be collected at the moment of the satellite overpass. A complete sounding survey, adequate for proper calibration of a remote bathymetric model, simultaneously with the satellite overpass, is logistically and economically impossible.

Time normalization of tidal fluctuations

To alleviate the problem of registering temporally-variant, tide-dependent water depth measurements with a somewhat static satellite image of the area, a time-normalization technique has been developed which can be used to rectify, or temporally normalize, sample water depth data with predicted tidal data. Three procedures are utilized in this technique:

(1) Construction of the tidal curve

The location of a reference tidal station closest to the area under study should be identified from appropriate tide tables. The predicted times and heights of the high and low waters at the reference tidal station on the date of satellite overpass can be found from the appropriate tide table and are used to construct a full tide curve by using the one-quarter, one-tenth rule which is presented in such tables (Figure 1)(NOS, 1987).

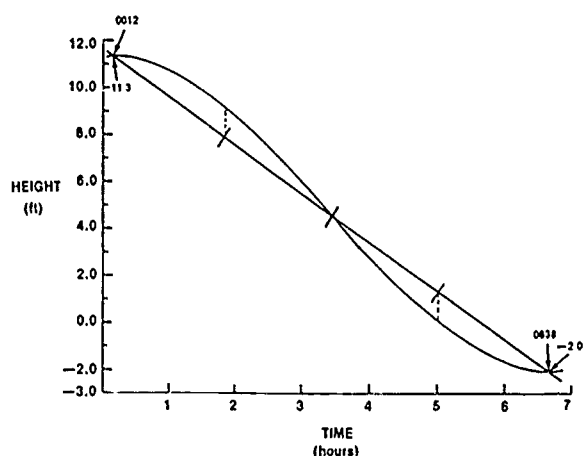


Figure 1. Tidal curve construction (from Tide Tables 1988, see References).

(2) Consideration of tidal subordinate stations

For those sampling sites which are not close to the reference station, subordinate stations which cover the sampling area should be selected from the tide table. For each subordinate station there will be differences for times and heights of high and low tides as compared to the reference station. Those differences are used to adjust the corresponding values of the reference station to construct modified working tidal curves for each sampling location (subordinate station).

(3) Normalization of tidal fluctuations

Use of this technique assumes that the satellite overpass of a study area is approximately instantaneous. This passage time can be obtained from the vendor of satellite data or retrieved from the satellite data tape (CCT) itself. With this time in GMT, the tidal height at the moment of satellite overpass can be read from the working tidal curve prepared. In the same way, tidal heights corresponding to each sampling time can be obtained and subtracted from the tidal height at the time of the satellite overpass. The resulting negative or positive values are the corresponding modifications for the time intervals between sample data collection and the satellite overpass. The sample water depths are then normalized to levels at the moment of satellite imaging by simply adding or subtracting the corresponding modifications.

To keep time systems consistent among data collections, tide records and satellite overpass, time system conversions (e.g. GMT to EST) must be done before normalizations are carried out.

Study areas and data collection

This study, a part of satellite remote bathymetry project, was conducted along the coast of Connecticut (CT), U.S.A. The reference station used was at New London, CT and two subordinate stations (data sampling sites) were chosen at Stonington (Fishers Island) and Noank (Mystic River entrance). Tide tables for the east coasts of North and South America for 1988 were used to obtain tidal predictions for the reference station and the time and height differences for the subordinate stations (NOS, 1987). Water depths were determined at each study site (subordinate stations) on May 9, May 13 and October 20, 1988, using an on-board recording fathometer. For each sounding, the time was recorded in Eastern Standard Time.

Data analysis

The experimental data were analyzed by comparing two data groups with and without normalization to evaluate the tidal effects on sample water depths in terms of varying locations and dates. Basic statistics, maximum absolute error, maximum relative error and mean relative error were used to represent the differences between raw sample data and normalized data. Results of the analyses are shown in the Table 1.

These results indicate that tidal fluctuations can result in a relative data error up to more than 30% in the study area and that tidal effects at Stonington were systematically different from those at Noank.

Conclusions and recommendations

Data normalization can significantly reduce the effect of tidal fluctuations on sample water depths and will be particularly applicable for water depth data in very shallow areas. Consequently, the accuracy of water depth models can be improved by using normalized sample data.

Tidal effects on data sampling vary with location and time, and it is apparent that predictions of water depths for a given model will be more comparable for different dates and locations using this normalization technique.

Tidal normalizing techniques would be more valid if actual tidal information for a specific reference station were to be used since direct tidal observations are more accurate than predicted values. Such information for North American coastal sites can be obtained from an agency such as the U.S. National Ocean Service.

The normalization technique of tidal fluctuations also makes it possible to use sample depth data collected either one day before or after the date of the satellite overpass by using an extended tide curve. This flexibility in collecting *in situ* depth calibration data makes possible, as well as practicable, the collection of larger samples of data, thus enhancing the statistical reliability of the remote bathymetric model.

Table 1. Errors between actual and normalized tidal heights at Noank and Stonington, CT

| Error | Dates of Water Depth Data Acquisition | | | | | |
|------------------------|---------------------------------------|------------|--------------|------------|------------------|------------|
| | May 9, 1988 | | May 13, 1988 | | October 20, 1988 | |
| | Noank | Stonington | Noank | Stonington | Noank | Stonington |
| Maximum absolute error | 0.21 ft | -0.36 ft | -0.56 ft | 1.18 ft | 0.05 ft | 0.76 ft |
| Maximum relative error | 3.7% | -5.0% | -6.8% | 30.9% | 1.5% | 17.4% |
| Mean relative error | 12.6% | -3.0% | -4.7% | 8.4% | 0.3% | 6.0% |

Acknowledgment

This research is supported in part by NOAA Grant NA-85AA-D-SG101 as Project R/O E-1 of the Connecticut Sea Grant Program. Scientific Contribution No.1284 of the Storrs Agricultural Experiment Station, The University of Connecticut.

References

Clark, R.K., T.H.Fay and C.L. Walker. "Bathymetry calculations with Landsat 4 TM imagery under a generalized ratio assumption", *Applied Optics*, Vol.26, No.19, pp 4036-4038, 1987.

Lyzenga, D.R., "Reflectance of a flat ocean in the limit of zero water depth", *Applied Optics*, Vol.16, No.2, pp 282-283, 1977.

Lyzenga, D.R., "Shallow water reflectance modeling with applications to remote sensing of ocean floor", *Proceedings of 13th International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, pp 583-602, 1979.

Lyzenga, D.R., "Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data", *International Journal of Remote Sensing*, Vol.2, No.1, pp 71-82, 1981.

National Ocean Service. "Tide Tables 1988 - East Coast of North and South America", U.S. Dept. Commerce, NOAA, Rockville, MD, pp 249, 1987.

Paredes, J.M. and R.E. Spero. "Water depth mapping from passive remote sensing data under a generalized ratio assumption", *Applied Optics* Vol.22, No.8, pp 1134-1135, 1983.

Polcyn, F.C. and I.J. Sattinger. "Water depth determination using remote sensing techniques", *Proceedings of 6th International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, Vol.2, pp 1017-1028, 1970.

Ross, D.A., "Introduction to Oceanography", 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 438 p, 1977.

Spitzer, D. and R. W. J. Dirks. "Bottom influence on the reflectance of the sea", *International Journal of Remote Sensing*, Vol.8, No.3, pp 279-290, 1987.

Spitzer, D. and R. W. J. Dirks. "Shallow water bathymetry and bottom classification by means of the Landsat and SPOT optical scanners", *SPIE Vol. 660 Earth Remote Sensing Using the Landsat Thematic Mapper and SPOT Sensor System*, pp 136-138, 1986.

A STUDY INTO THE RESPONSES OF THE NOAA-n AVHRR REFLECTIVE CHANNELS OVER WATER TARGETS

Sonia C. Gallegos, Thomas I. Gray, and Melba M. Crawford

Center for Space Research
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

The capabilities of the AVHRR reflective channel data to monitor phytoplankton distributions at the ocean surface are investigated. To utilize this data, a procedure to correct for the effects of atmospheric water vapor and pathlengths associated with those pixels away from the subsatellite point was developed. Data obtained with this methodology is compared with near simultaneous CZCS data and *in-situ* observations. The AVHRR and the CZCS data were segmented into homogeneous regions using a mathematical classification technique and then compared to determine similarity of spatio-temporal features.

Keywords: AVHRR reflective, CZCS, phytoplankton, red tide, atmospheric correction.

INTRODUCTION

Since the CZCS stopped broadcasting in 1984, there has been a great need for a spaceborne instrument which could provide insight into the phytoplankton distribution at the ocean surface. The data of the Advanced Very High Resolution Radiometer (AVHRR) from the TIROS-N/NOAA-n satellite series can partially fulfill this requirement. This instrument was not specifically designed to observe phytoplankton. In fact, its primary purpose is to monitor hydrological and meteorological processes for environmental studies. Although the thermal bands of the AVHRR are presently utilized to determine sea surface temperature (SST) (McClain et al, 1981), the reflective bands have limited use in oceanographic work (Gallegos, 1984). Conversely, the reflective channels have been widely employed in terrestrial vegetation surveys since 1981 (Gray and McCrary, 1981).

This study was undertaken to investigate the potential use of AVHRR data in monitoring water targets. An atmospheric water vapor correction was developed and applied to the radiometrically corrected data. A factor was also proposed to adjust response according to pixel position in the swath. The processed AVHRR imagery then was segmented using a supervised classification technique and compared to nearly concurrent classified CZCS imagery. In addition, locations of red tide in the AVHRR imagery were compared to *in situ* data for a 1983 red tide incident off the west coast of Florida.

METHODS

I. Correction factors for AVHRR data

Digital data from High Resolution Picture Transmissions (HRPT) (Figure 1) and near-coincident CZCS pigment calibrated data (Figure 2) were acquired. The AVHRR data for the area of the eastern Gulf of Mexico was extracted, calibrated, and edited. In terrestrial vegetation studies, the AVHRR data is typically displayed in 8-bit words, obtained by dropping the two least significant bits (LSB) of the original 10-bit data. Unfortunately, this is where the information for water targets is stored. This is done to fit the requirements of most imaging systems. The editing procedures used in this study permit the utilization of the full 10-bit resolution of the AVHRR data sets in the low albedoes pertaining to water targets and compress the high albedo values associated with clouds into narrow intervals. Specifically, albedo values greater than 22.5 are compressed into ranges of 5% albedoes. Values below this threshold are multiplied by 10. A similar procedure is used for the thermal channels. The threshold value for the emissive channels is -4.16 °C, and the data is compressed into 16 °C intervals. The resulting data exhibits a sensitivity of 0.1% in the reflective channels and increments of 0.25 °C for the thermal bands.

The atmospheric adjustments were made under the assumptions that 1) the available solar energy is constant, 2) there was no significant underwater reflectance contained in either of the reflective channels, 3) Mie Rayleigh and aerosols were known computable constants and 4) water vapor, principal attenuator at these wavelengths, can be estimated from AVHRR data. To obtain the correction factor for water vapor a piecewise linear regression model was developed utilizing saturation precipitable water information from 72 upper air soundings and the responses of the reflective channels 1 (585 nm - 685) nm and 2 (713 nm - 986 nm) and the emissive channel 4 (10.362 nm - 11.299 nm) of the AVHRR.

The NOAA-n satellite crossing times over the Gulf of Mexico did not coincide with the sounding times. Therefore the location of the air mass at the time of the satellite pass had to be extrapolated based on local sounding information on wind velocity and direction. Radiances for the reflective and emissive channels were then extracted at that position. These data were input in various regression models. The model which fit the data the best ($r=0.83$), had "adjusted precipitable water" as the dependent variable and the emitances of Channel 4 as the independent variable. "Adjusted" is defined here to be the integrated saturation precipitable water values multiplied by the response of Channel 1. The predicted values were used as an index that represents a measure of the atmospheric water vapor at the time of the satellite pass.

There are significant distortions in the data of the AVHRR because of its wide instantaneous field of view (IFOV) which spans 56° on either side of nadir. A correction procedure was developed in this study to correct for this distortion. It is based on the concept that the footprint of the instrument increases and changes shape from a circle to an ellipse as it moves from nadir toward the edge of the swath. The increased area of the footprint results in increased radiating surfaces. Adjusted reflectance estimates with reference to pixel position in the swath, require determination of the pixel size with respect to spacecraft height above the earth, earth curvature and changes in the nadir angle as it moves away from nadir. The proposed normalizing factor is the ratio of the nadir area to the area of the given pixel. The combined atmospheric and geometric correction procedure is applied to each pixel.

II. Validation of AVHRR corrected data

Comparison of the capability of different satellite systems to monitor a target is difficult. One approach is to classify imagery from an individual system and to then compare the resulting classified images. In this study, a supervised piecewise linear classification algorithm was trained for CZCS pigment calibrated data and then for AVHRR data using observations from geographic locations where red tide was reported. The technique which is based on a seniority logic committee approach (Lee and Richards, 1984) was modified and implemented on a HP 900/835 by Khazenie (1987). The method segments the multispectral

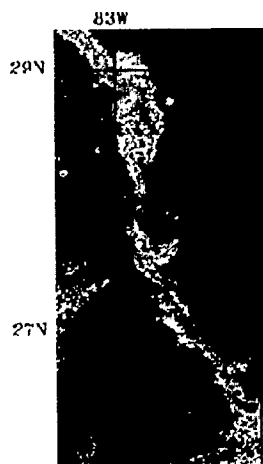


Figure 1. AVHRR corrected reflective data (Channels 1 and 2) October 26, 1983

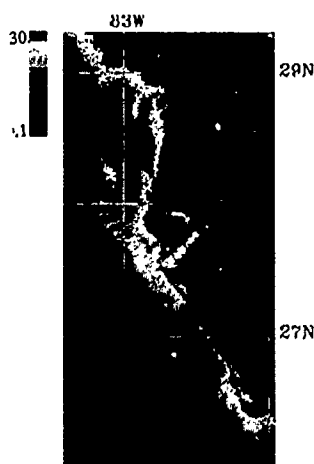


Figure 2. CZCS Pigment calibrated data October 27, 1983

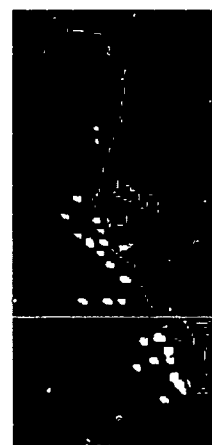


Figure 3. Red tide sightings off the Florida west coast for the month of October 1983

patterns in a practical, simple manner, has an adaptable committee size, and is linear with respect to the number of classes. Training is computational intensive, both in time and memory requirements. However, once trained, the algorithm can be run on very large images with no difficulty. Results of the classification procedure are shown in Figures 4a - 4c.

CONCLUSIONS AND RECOMMENDATIONS

The new editing procedures improved the sensitivity of the reflective channels response over water targets. The increased sensitivity resulted in small fluctuations in the data over areas where the signal appears flat in the 8-bit processed data. The addition of the combined correction algorithm permitted some normalization of the spectral response of the data obtained on different dates during the study period.

Investigation of the reflective channels under conditions of clear and turbid atmospheres by Duggin et al (1982) indicates that of the 2048 pixels contained in an AVHRR swath, only the central 512 pixels (256 on each of nadir) are free from excessive nadir angle distortion. The results obtained in the current study suggest that the range of acceptable data could be extended to at least 512 pixels on either side of nadir. Pixels exhibiting pixel numbers greater than approximately 1600 were not utilized in this study because they were contaminated by sunglint. Several normalization procedures were investigated to handle this problem. None of the approaches provided satisfactory results and were not included here.

The classified images indicate that there were spatio-temporal features common to both the AVHRR and CZCS data. These features coincided with chlorophyll concentrations greater than 1.5 mg m^{-3} . In some regions where the AVHRR and CZCS signatures exhibited the same general outline but failed to coincide in a pixel by pixel basis, the differences appear to be related to environmental changes occurring between acquisition times. The AVHRR data presented here was collected on October 26, 1983 at about 14:30 Local solar time (LST) and the CZCS on October 27, 1983 at approximately 12:00 LST.

Comparisons between the AVHRR scene (Figure 1) and the ground truth locations of the 1983 red tide incident off Florida (Figure 3), indicate that all of the ground truth stations reporting sightings of red tide (white dots) occurred within the area of the red tide signature identified by the AVHRR. Results from this investigation suggest that the AVHRR reflective channel data can be used to monitor certain concentrations of phytoplankton, especially those above 1.5 mg m^{-3} . This instrument can be a valuable tool in tracking red tide blooms.

The results of the normalization adjustments are promising, especially over clear and semi-clear data. These techniques do not seem to work as well for areas with amounts of precipitable water greater than 6.0 g kg^{-1} . It is recommended that these methods be applied to data from other oceanographic sites and tested over longer periods of time for full validation.

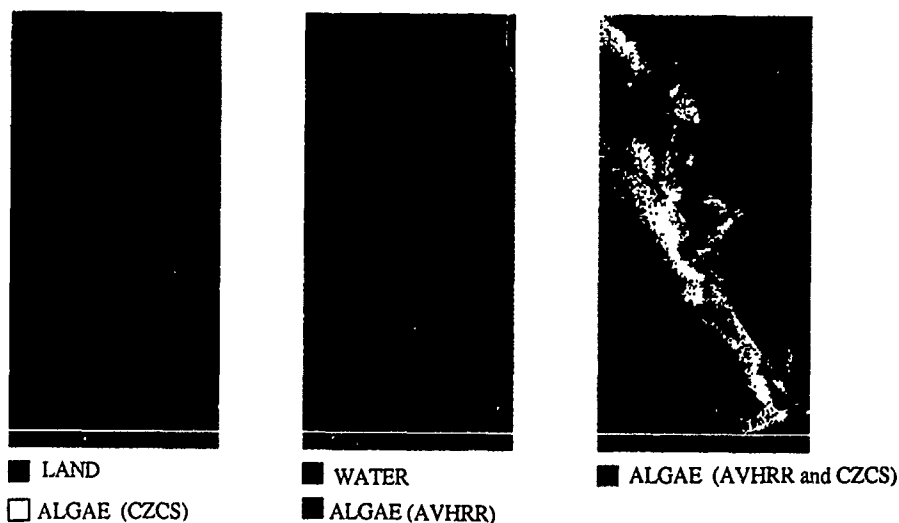


Figure 4 (a) Linearly classified CZCS data for October 27, 1983 (b) Linearly classified AVHRR data for October 26, 1983 (c) Differences between CZCS and AVHRR classified data

ACKNOWLEDGEMENTS

The authors would like to thank the following individuals and institutions who generously contributed to this investigation. The Florida Department of Natural Resources and in particular Ms. Beverly Roberts for making available field records for the 1983 red tide event off Florida; Mr. Dennis Clark of NOAA/Ocean Science Division for processing the pigment calibrated CZCS data; Ms. Kyungsook Kim of the Center for Space Research for helping with the data classification, the students at the Center for Space Research who contributed in one way or another to this study.

This project was supported by the Texas Advanced Research Technology Project (TATRP) under contract # 14-9710, the Texas A & M Sea Grant Program - Grant # 26-3904 and the Center for Space Research of the University of Texas at Austin.

REFERENCES

- Duggin, M.J., D.Piwinski, V. Whitehead and G. Ryland, Evaluation of NOAA AVHRR data for crop assessment. *Appl. Opt.* 21, 1873-1875, 1982.
- Gallegos, S.C., R.S. Nerem, T.I. Gray and M.R. Helfert, Vegetative responses from a Great Barrier Reef surface water feature detected by Space Shuttle photography. Technical paper 1984 ASP-ACSM Fall Convention, San Antonio, Texas, 1984.
- Gray, T.I. and D.G. McCrary, Meteorological satellite data - a tool to describe the health of the world's agriculture, EW-N1-04042, JSC-17112, AgRISTARS Early Warning and Crop Condition Assessment Project, U.S. Department of Commerce NOAA, Houston, Texas, 1981.
- Khazenie, N, Contextual Classification of remotely sensed data based on a spatial-temporal correlation model, Ph.D Dissertation. University of Texas at Austin, 1987.
- Lee, T., and J.A. Richards, Piecewise linear classification using seniority logic committee methods with applications to remote sensing, *Pattern Recognition* 17, 453-464, 1984.
- McClain, P.E., Multiple atmospheric-window techniques for satellite-derived sea surface temperatures, In: *Oceanography from Space* Ed: Gower, Plenum Publishing Corp., 73-85 1981.

OVERVIEW AND HIGHLIGHTS FROM THE NORCSEX'88: A PRE-LAUNCH ERS-1 EXPERIMENT

Johnny A. Johannessen

Ola M. Johannessen

Robert A. Shuchman

Nansen Remote Sensing Center
Bergen, Norway

Nansen Remote Sens. Center
Geophysical Institute, Univ. of Bergen, Norway

Radar Science Laboratory
Environmental Research Institute
of Michigan, Ann Arbor, MI, USA

ABSTRACT

This paper provides an overview of the pre-launch ERS-1 NORwegian Continental Shelf EXperiment (NORCSEX) carried out in March 1988 off the west coast of Norway centered at 64°30' N. The overall objective was to simultaneously measure marine variables such as near surface wind, waves and current and their interaction during winter conditions by integrated use of remote sensing and in-situ data collection. The capability of the ERS-1 type active microwave instruments to sense these variables was investigated by combined use of an aircraft X- and C-band SAR, the GEOSAT radar altimeter and a ship mounted X-, C- and L-band scatterometer. In-situ collection of oceanographic and meteorological data was simultaneously obtained from research vessels and moorings. The ongoing analyses are providing interesting results especially from: a) the SAR imaging of ocean features; b) the SAR wave studies; c) the scatterometer versus wind speed relationship; d) the multisensor SAR-radar altimeter comparison; and e) the SAR-scatterometer comparison.

ERS-1, WIND-WAVE-CURRENT, DATA BASE, MULTI-SENSOR ANALYSIS

1. Introduction

The first European Earth Resource Satellite (ERS-1) will be launched into polar orbit during the winter of 1990/91. In order to demonstrate the potential capability of the active microwave sensors onboard this ESA ERS-1 satellite to measure near surface wind, ocean waves and current and their interaction, the NORCSEX'88 pre-launch ERS-1 experiment was carried out off the west coast of Norway (Figure 1) in March 1988. For this particular and many other coastal regions subject to extensive cloud cover as well as a weakening in the surface temperature gradient across the ocean circulation features during seasonal warming, the ERS-1 satellite may provide the first platform for regular observations of these marine variables.

ERS-1 type remote sensing data were collected by a CV-580 aircraft X- and C-band SAR, the GEOSAT radar altimeter and a ship mounted X-, C- and L-band

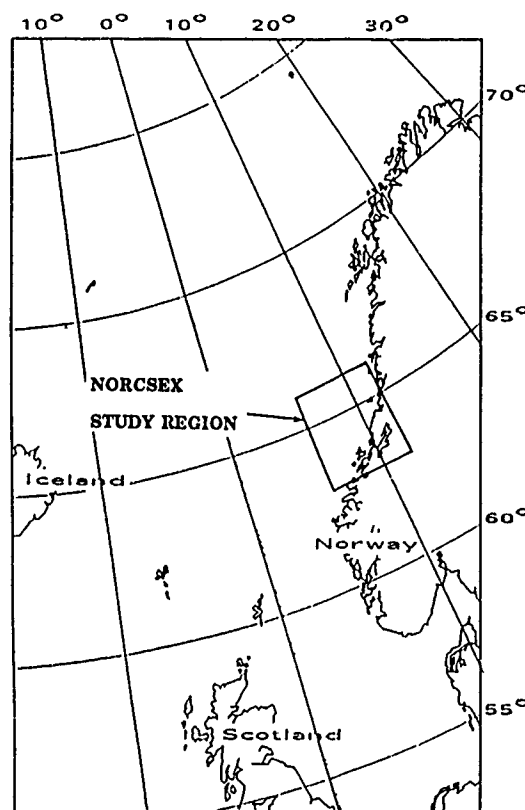


FIGURE 1. THE NORCSEX STUDY REGION

scatterometer. In addition a multi-platform (satellites, research vessels, moorings) and multi-sensor sampling strategy provided observations of meteorological and oceanographic "ground truthed" data of near surface wind speed, atmospheric boundary layer stability, ocean waves and current. The data collection overview is summarized in Table 1.

The primary in-situ instrumentation were the directional wave buoys (Wavescan) providing wave spectral estimates, the ship- and buoy mounted hot film and mini-cup anemometers for atmospheric boundary

TABLE 1. DATA COLLECTION OVERVIEW

| PLATFORM INSTRUMENT | VARIABLES | | |
|-----------------------|-----------|-------|---------|
| | WIND | WAVE | CURRENT |
| GEOSAT Altimeter | S | Hs | T,E |
| DMSP SSM/I | S | | F,E |
| NOAA AVHRR | | | |
| CONVAIR 580 SAR (X,C) | + | L,D | F,E |
| WAVESCAN Buoy | S,D | H,L,P | S,D |
| Current Meters | | | |
| Turbulent Flux Buoy | S,D | | S,D |
| CODAR | | | |
| HÅKON MOSBY - ADCP | | | S,D |
| Turbulence | S,D | | |
| Scatterometer | + | | |
| Ship Radar | | + | |

S: speed
H: wave height
F: fronts
L: wavelength
Hs: significant waveheight

D: direction
T: ocean topography
E: eddies
+: some info
P: period

layer wind speed and turbulence measurements, the Coastal Ocean Doppler Aperture Radar (CODAR) for surface current measurements, anchored moorings for surface and subsurface current measurements, digitizing of the ship X-band radar data attempted for wave studies along the ship track, and the Acoustic Doppler Current Profiler for absolute current measurements along the ship track. In addition the thermohaline structure and sea surface temperature were continuously measured with a towed, undulating SeaSoar and a ship mounted termistor. Overall the intensive data collection during the one month field program provide a unique data base as schematically illustrated in Figure 2.

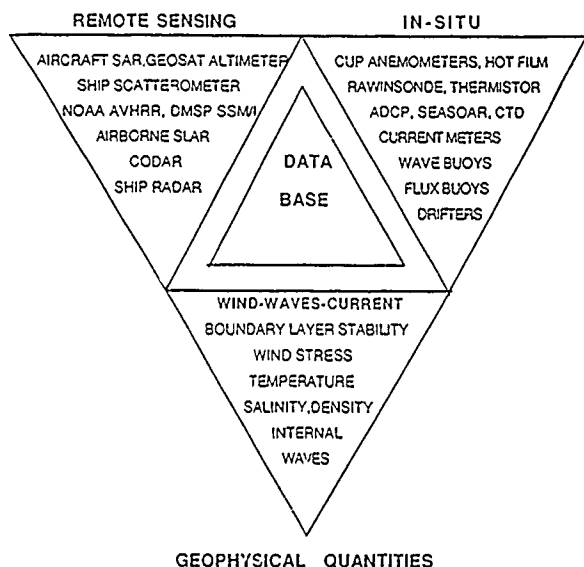


FIGURE 2. SCHEMATIC DATA BASE

In this overview paper the status of the remote sensing observations and presentation of a few highlight results are reported in section 2 subsequently followed by a summary in section 3.

2. Observations

The characteristics of the winter environmental conditions during the 30 day NORCSEX'88 field program obtained from in-situ observations and supported by the NOAA satellite AVHRR data were; a) passages of three storms with maximum 10 minutes average wind speed of 25 m/s; b) a prevailing unstable boundary layer stratification from -2° to -5° C; c) pre-dominantly southwesterly waves with a mean wave height ranging from 1m to 6m and with maximum recorded waveheight of 10.5m; and d) the existence of a well defined meandering frontal boundary of $3-4^{\circ}$ C between the northward flowing Norwegian Coastal Current and Atlantic water accompanied by a current shear reaching $10^{-4} s^{-1}$. The status of the simulated ERS-1 type data collection is reported in the next paragraphs.

a. *Radar Altimeter.* A location map of the GEOSAT ground tracks for the 17 day repeat cycle covering the investigation area during the field program is shown in Figure 3. The mooring site for the four Wavescan Buoys located in cross-over arcs are indicated with circles. Both

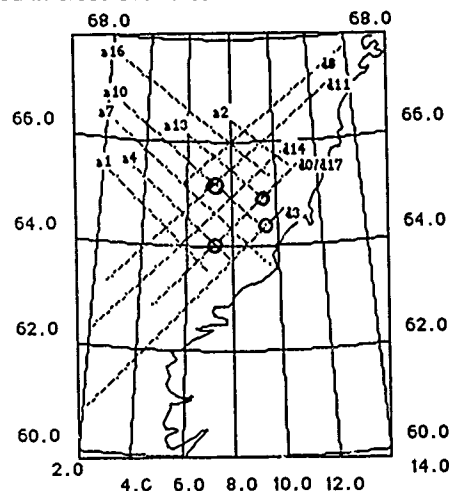
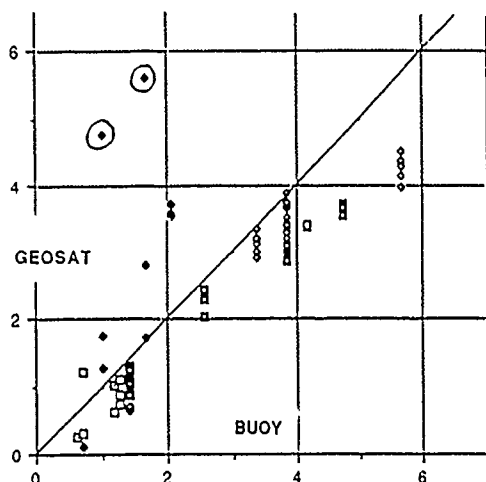


FIGURE 3. GEOSAT COVERAGE

the wind and significant waveheight provided by the altimeter are in favourable agreement with weathermaps and in-situ wind speed and directional wave measurements. The altimeter seems to underestimate the significant waveheight for values larger than 1m (Figure 4). Likewise the altimeter underestimate high wind events, while rapid wind shifts are smeared out.

From repeated track analysis the altimeter height measurements provide estimates of time varying mesoscale circulation pattern with maximum surface elevation c. 20 m over a horizontal distance of 40 km. This yields geostrophic current of .30 - .40 m/s. These patterns are mostly located on the wide continental shelf in water of 300 m depth. The current magnitude agrees with direct current measurements as well as location of mesoscale eddies seen from satellite AVHRR imagery. Provided the geoid is accurately known or that repeated track analysis with sufficient record length is available the radar altimeter is capable to provide useful information of the mesoscale circulation pattern along the Norwegian Continental Shelf.

FIGURE 4. GEOSAT - BUOY COMPARISON



More specific results of the altimeter investigation during NORCSEX'88 are reported in; Session L2 (NORCSEX 1) papers 5, 7 and 8; and Session L3 (NORCSEX 2) papers 5 and 7.

b. Scatterometer. The ship mounted X-, C- and L-band scatterometer imaged the ocean surface over a range of incidence angles from 20° to 80° . Data were collected continuously for 15 days at winds ranging from calm to 25 m/s primarily at unstable stratification in the boundary layer. Under conditions when the wind and the wave field in the scatterometer footprint are undisturbed by the presence of the ship these "ground truthed" remote sensing data collected together with near surface wind speed (and wind stress) can be directly used for the study of the relationship between the near surface wind speed and radar backscattering cross section. In Figure 5 this relationship is shown for selected data when the ship crossed a sharp wind front and the wind speed increased from about 2 m/s to 12 m/s in a 1/2 hour. The

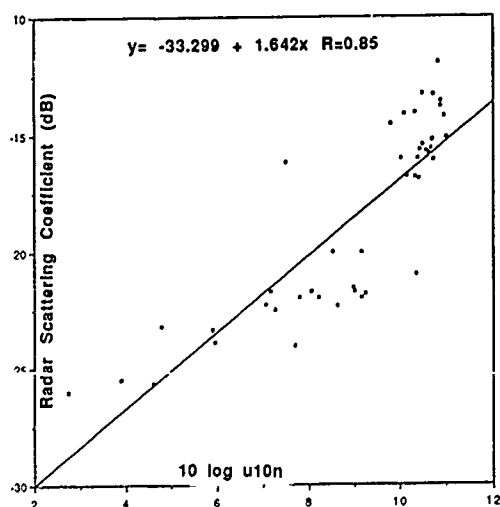


FIGURE 5. BACKSCATTER VERSUS WIND SPEED

corresponding increase in backscatter (C-band, hh polarization) is seen to exceed 10 dB suggesting a ratio of about one between wind speed increase (logarithmic) and backscatter increase (dB).

The results of the ship mounted scatterometer studies are more extensively reported in; Session L2 (NORCSEX 1) paper 2; and Session L3 (NORCSEX 2) papers 1 and 2.

c. SAR. The aircraft SAR investigation focused on wave studies and ocean feature studies. In addition a ship and ship wake study using SAR was carried out (to be presented under Session L1 (SAR/Ocean) - Ship wake observations). In total SAR data was collected for 30 hours including multi-sided flight pattern over Wavescan Buoys, mosaic flight pattern over the ships, the moorings and the CODAR swath, and about 400 km of GEOSAT radar altimeter underflight.

The sea state present during the SAR wave flights are summarized in Figure 6, and shows that range and azimuth travelling waves were encountered for a variety of wave heights ranging from 1m to 6m. Likewise the flights were carried out for a significant range in the near surface wind speed. The use of the wave spectra obtained from the Wavescan Buoys as well as the significant waveheight obtained from the radar altimeter are important for the SAR wave studies especially focusing on the analysis of azimuth smearing, estimates of significant waveheight spectra, and the 180° directional ambiguity problem. Specific SAR ocean wave investigations are reported in; Session L3 (NORCSEX 2) papers 6, 7 and 8.

Despite the fact that Seasat demonstrated that the L-band SAR can observe detailed structure in the mesoscale circulation field, SAR has not been extensively used for studies of the ocean circulation processes. SAR backscatter maps cannot be related directly to geophysical quantities, largely because of the lack of in-situ ocean circulation experiments. Consequently the SAR mosaic flight experiment was conducted such that the real time processed SAR data was used to guide the ship in order to obtain near simultaneously in-situ measurements of significant structures. The ship scatterometer data can also be directly compared with the simultaneously obtained aircraft SAR data since resolution and imaging mechanisms are approximately the same.

The SAR imagery contain several distinct features

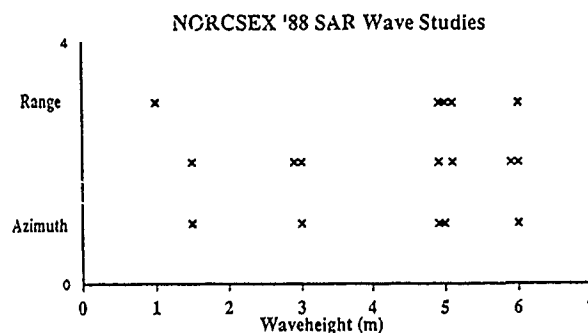


FIGURE 6. SEA STATE DURING SAR FLIGHTS

modulated by surface current shear zones and wind fronts. The different appearances of these features in the SAR imagery may be utilized in a SAR classification scheme. Ocean current modulated features are typically seen as narrow zones of intense backscatter surrounded by uniform weaker backscatter. Rapid wind shifts on the other hand are clearly detected by the SAR as sudden changes in backscatter from one intensity level to another. In turn such zones may have impact on the sea state leading to significant changes in wave direction and wave height. Evidence of wave refraction by current shear are also found. The multisensor SAR - radar altimeter investigation is important in this analysis. Furthermore since the boundary layer stratification is unstable the turbulent convective cell circulation generates wind modulated pattern on the surface that may provide estimates of the wind direction.

Further details of the analysis of the SAR imaging capabilities of ocean circulation features and atmospheric wind effects are reported in; Session L2 (NORCSEX 1) papers 2, 3; and Session L3 (NORCSEX 2) papers 1, 2, and 3.

4. Summary

The wind-wave-current data collected during the multiplatform and multisensor NORCSEX'88 pre-launch ERS-1 investigation have generated a unique data base needed to test, improve and develop algorithms for deriving measurements of these marine variables from ERS-1 and other future satellites. The aim is to complete the ongoing analysis of the pre-launch data (the next 15 papers in the NORCSEX special session) in time before the launch of ERS-1. The advantage of reaching final results before launch will in turn improve the conduct of the post-launch ERS-1 experiment as well as improving the interpretation of the ERS-1 data. This is required in order to prepare for operational use of the data.

SAR DETECTION OF MESOSCALE OCEAN CIRCULATION FEATURES DURING NORCSEX '88

Robert A. Shuchman and David R. Lyzenga
Radar Science Laboratory
Environmental Research Institute of Michigan
Ann Arbor, MI 48107 USA

Johnny A. Johannessen and Ola Johannessen
Nansen Remote Sensing Center
Bergen, Norway

Charles E. Livingstone
Canada Centre for Remote Sensing
Ottawa, Ontario K1A 0Y7

ABSTRACT

During the Norwegian Continental Shelf Experiment (NORCSEX) simultaneous C- and X-band synthetic aperture radar (SAR) data were collected over fronts and eddies associated with the Norwegian coastal current. The fronts and eddies were well documented with respect to their meteorological and oceanographic character. Both a current and wind induced frontal feature were observed by the SAR. Under unstable low wind conditions, the C-band SAR observed current shear on the order of 10^{-3} sec^{-1} . The wind front was separable from the SAR observed current shear by noting the step-like character of the boundary on the imagery.

1. INTRODUCTION

A wind-wave-current field investigation using active microwave sensors was conducted on the Norwegian Continental Shelf in March of 1988 (NORCSEX '88). One of the primary objectives of NORCSEX which was a prelaunch European Space Agency ERS-1 investigation was to investigate the ability of a C-band synthetic aperture radar (SAR) to detect mesoscale ocean circulation features.

During NORCSEX simultaneous X- and C-band SAR data were collected over fronts and eddies associated with the Norwegian coastal current. The fronts and eddies were well documented with respect to their oceanographic and meteorologic character. For example, wind speed and direction, wind stress, drag coefficients, surface currents, gravity waves and sea surface temperature were continuously measured.

SAR detection of mesoscale circulation features is a function of: atmospheric stability as related to sea surface temperature variations, enhanced wave breaking at the Bragg scale, slope and refraction changes of the dominant gravity wave, wind discontinuity, direct interaction of Bragg waves with currents, and redistribution of surfactant material. The NORCSEX data is being utilized to quantify each of the imaging mechanisms discussed above. Initial analysis indicates that a single mechanism does not dominate the observed SAR signature, but rather depends on a diverse set of oceanographic and meteorologic conditions.

The SAR observed ocean fronts to be discussed in detail in this paper are divided into two broad categories: 1) wind fronts with a backscatter change of 7-10 dB; and 2) current fronts with a backscatter change of 2-4 dB. In this paper we will first describe the SAR data collected during NORCSEX. This

will be followed by a presentation of the current and wind induced fronts observed. The classification of the SAR observed fronts into the two categories will then be supported through comparison with an ERIM SAR backscatter simulation model for the ocean.

2. DATA SETS

The aircraft CV-580 C- and X-band SAR observations were complemented with synoptic *in situ* measurements of boundary layer oceanic and atmospheric variables by instruments mounted on research vessels and moored platforms.

The oceanographic sea truth was obtained by the use of a ship mounted thermistor (sea surface temperature), a towed undulating SeaSoar (salinity and temperature from the surface to 250 m), and the ship mounted 150 KHz Acoustic Doppler Current Profiler (ADCP). The ADCP provides a measure of absolute current every 5 m from the surface to near the ocean bottom. Moored platforms included current meters and pitch and roll buoys with meteorological packages.

The meteorological measurements and data assimilation were conducted during the entire NORCSEX field investigation period. Time series of surface layer meteorological data from ship mounted sensors and profilers of temperature, humidity and vector wind from rawinsondes were obtained. Surface layer wind fluxes (i.e., drag coefficients) were obtained from ship mounted hot-film and the use of miniature cups.

In addition to the SAR data, other remote sensing data included the ship L-, C-, and X-band scatterometer, German CODAR, the NOAA satellite Advance Very High Resolution Radiometer (AVHRR) imagery, the GEOSAT satellite altimeter, and DMSP satellite Special Scanning Microwave Imager (SSM/I).

Table 1 is a summary of the SAR flights conducted during NORCSEX. Note from the table that four mosaic SAR flights were carried out for mesoscale ocean feature studies. The SAR system parameters during these flights are summarized in Table 2 along with the system parameters for the ship mounted L-, C-, and X-band scatterometer. A complete description of the CV-580 C- and X-band SAR system is given in Ref. [1].

3. SAR OBSERVATIONS

For SAR and scatterometer imaging the ocean over a range of incidence angles from approximately 20° to 80° , the radar backscatter is principally caused by a resonant mechanism, called Bragg scattering where the radar waves (and energy) are scattered by surface waves of approximately the same wavelength [Ref. 2]. These short surface waves are modulated by air-sea interaction processes, by long wave-short wave

TABLE 1. CV-580 SAR FLIGHT SUMMARY

| DATE MARCH 1988 | FLIGHT PATTERN | INVESTIGATION | COMPLEMENTARY DATA |
|--------------------|-----------------------|------------------------------|---|
| 11 | Multisided | Waves | Wavescan, GEOSAT D0 and A1, Hakon Mosby |
| 13 | Mosaic | Ocean features | Hakon Mosby, CODAR, moorings. |
| 14 | Multisided/ Mosaic | Ocean features/ waves | Hakon Mosby, CODAR, GEOSAT D3 and A4, Wavescan and moorings High quality NOAA AVHRR |
| 17 | Mosaic | Ocean features | Hakon Mosby, GEOSAT A7, CODAR Low quality NOAA AVHRR |
| 20 | Multisided/ Mosaic | Waves | GEOSAT A10, Wavescan, CODAR |
| 21 | Multisided/ Mosaic | Ship wake/ Ocean features | Hakon Mosby, CODAR, Moorings |

interaction, by the presence of surface films and by oceanic internal processes causing sufficient variability that can be seen on the high resolution (10 - 30 cm) SAR imagery. In general, no single modulation mechanism dominates the observed SAR backscatter signature but rather depends on a variety of oceanographic and meteorological conditions.

TABLE 2. CV-580 SAR and Ship Scatterometer System Parameters

| Parameter | SAR | | Scatterometer | | |
|-----------------------------|--------|--------|---------------|--------|--------|
| | X-band | C-band | X band | C band | L-band |
| Frequency (GHz) | 9.35 | 5.3 | 9.50 | 5.25 | 1.50 |
| Wavelength (cm) | 3.2 | 5.6 | 3.2 | 5.6 | 20.0 |
| Polarization | VV | VV | VV/VH | VV/HH | VV/HH |
| Aircraft altitude (km) | 7 | 7 | | | |
| Aircraft speed (m/s) | 125 | 125 | | | |
| Scatterometer height (m) | | | 10 | 10 | 10 |
| Ship speed (m/s) | | | 4 | 4 | 4 |
| Incidence angle (degrees) | 45-85° | 45-85° | 20-70° | 20-70° | 20-70° |
| Pulse length (microsec) | | | | | |
| Bandwidth (MHz) | 11.5 | 11.5 | 575 | 525 | 350 |
| Azimuth beam width (deg) | 4.2° | 4.2° | 2.6° | 2.6° | 2.6° |
| Ground range resolution (m) | 15 | 15 | 0.3 | 0.3 | 0.3 |
| Azimuth resolution (m) | 7.5 | 7.5 | 0.3 | 0.3 | 0.3 |
| Look direction | L/R | L/R | Starboard | | |

* For scatterometer at 40° incidence angle

Figures 1 and 2 are mosaics generated from the CV-580 C-band vertical transmit and receive polarization (VV) data for 13 and 17 March 1988 respectively. The wind was light (3-4 m/s) out of the southwest (230°T) on the 13th of March. The frontal feature (see arrow on Fig. 1) observed on the 13th data is a result of current shear. On the 17th, the wind varied across the mosaic, fluctuating from 2 to 13 m/sec. The wind direction for this case was northwest (i.e., 290°T). The prominent front (see arrow on Fig. 2) on the 17th data is due to changes in the surface wind. Examination of both the mosaics confirms the existence of a complicated mesoscale circulation pattern associated with meander growth, eddy formation, and intrusions of Atlantic water toward the coast.

During the NORCEX field investigation, typical winter conditions characterized the oceanic and atmospheric environment on the continental shelf. A prevailing unstable boundary layer varying from -2 to -5°C existed, which implies larger wind speed near the sea surface than for stable conditions at the same wind speed. A swell field with a minimum long wave height of 2-3 m was always present along with a well defined ocean frontal boundary of 3-4°C between the coastal water and Atlantic water.

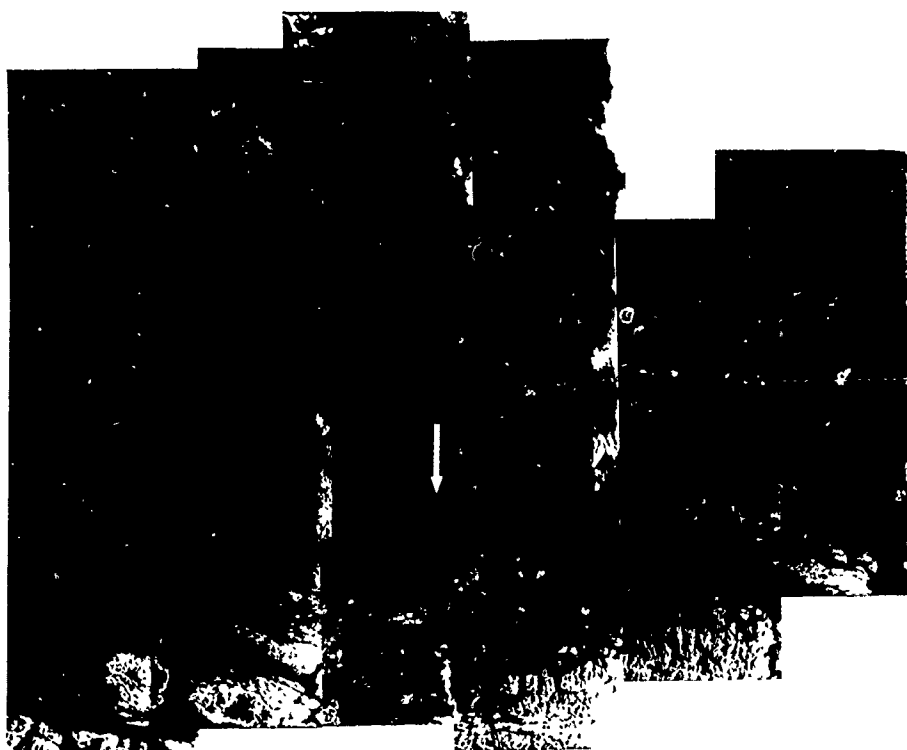
The distinct frontal features imaged by the aircraft SAR on 13 and 17 March are enlarged and shown in Figures 3 and 4, respectively. The SAR front A (Fig. 3) is characterized by a narrow, bright line of increased backscatter curving slightly in range direction with equal darker zones of less intense backscatter on both sides. Similar structures interpreted to be modulated by mesoscale surface current pattern were frequently seen in Seasat SAR images of the Gulf Stream [Ref. 3].

In contrast to SAR front A, front B (Fig. 4), mostly oriented in azimuth direction, is characterized by a rapid change from dark to brighter SAR backscatter return (i.e., greater radar backscatter). The corresponding profiles of the backscatter variations across these frontal features are shown for comparison in Figure 5. The backscatter of the bright line appears as a bump with a width of about 1 km exceeding the surroundings by about 2 dB. On the other hand, a step like profile with an 8 dB increase over 1 km is encountered across SAR front B. The different appearances of these two features immediately suggest that their nature of origin are likely to be different. This is supported by the *in situ* measurements.

R/V Hakon Mosby was directed from real-time processing of the SAR data onboard the aircraft, thus providing necessary *in situ* observations of the ocean surface and atmospheric boundary layer variables across the SAR detected frontal features. The characteristics of these variables representing the conditions at the time of the SAR flights are quantified in Table 3. The major differences occur in wind speed and ocean current shear. The other important geophysical quantities and sources for SAR backscatter modulation such as the magnitude of the unstable boundary layer conditions, the surface ocean temperature front and the long wave field remains nearly unchanged. Consequently, the primary modulation mechanisms generating these two features are interpreted to be, respectively, interaction of Bragg waves with a current shear and adjusting Bragg waves to a rapid wind increase. These interpretations are discussed below.

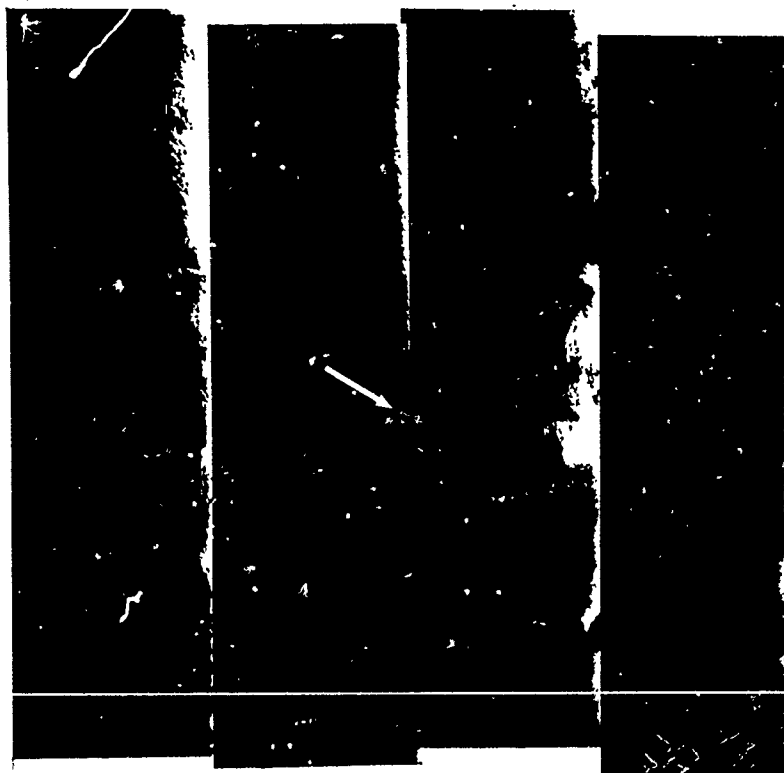
TABLE 3. Geophysical Quantities Contributing to the SAR Images Together With SAR System Parameters

| | 13 09-13 | 17 07-11 | March hours |
|---|------------------|------------------|----------------|
| T E M P E R A T U R E | | | |
| SEA | 5-7°C | 5-8.5°C | |
| AIR | 2°C | 3°C | |
| FRONT | 2°/7km | 3°/15 km | |
| STABILITY | Unstable | Unstable | |
| S W E L L | | | |
| H1/3 PERIOD | 3m | 2-4m | |
| DIRECTION | 11 sec 300° | 12 sec 300° | |
| W I N D | | | |
| MAX (m/s) | 6 NW | 11 NW | |
| MIN (m/s) | 4 W | 2 S | |
| U* | 0.25 m/s | 0.40 m/s | |
| C U R R E N T | | | |
| MAX (m/s) | 0.36 E | 0.50 NW | |
| $\partial U/\partial x + \partial V/\partial y$ | 10 ⁻⁴ | 10 ⁻⁴ | |
| $\partial V/\partial y - \partial U/\partial x$ | 10 ⁻³ | 10 ⁻⁴ | |



NORCSEX '88
CCRS CV 580 SAR Data
C Band (VV)
13 March 1988
0900 1300 UT

Figure 1. C-Band SAR 13 March 1988 Mosaic Showing Frontal Features



NORCSEX '88
CCRS CV-580 SAR Data
C-Band (VV)
17 March 1988
0800 1100 UT

Figure 2. C-Band SAR 17 March 1988 Mosaic Showing Frontal Features



Figure 3. Enlargement of the 13 March 1988 C-Band (VV) SAR Data Showing the Current Shear Induced Ocean Front

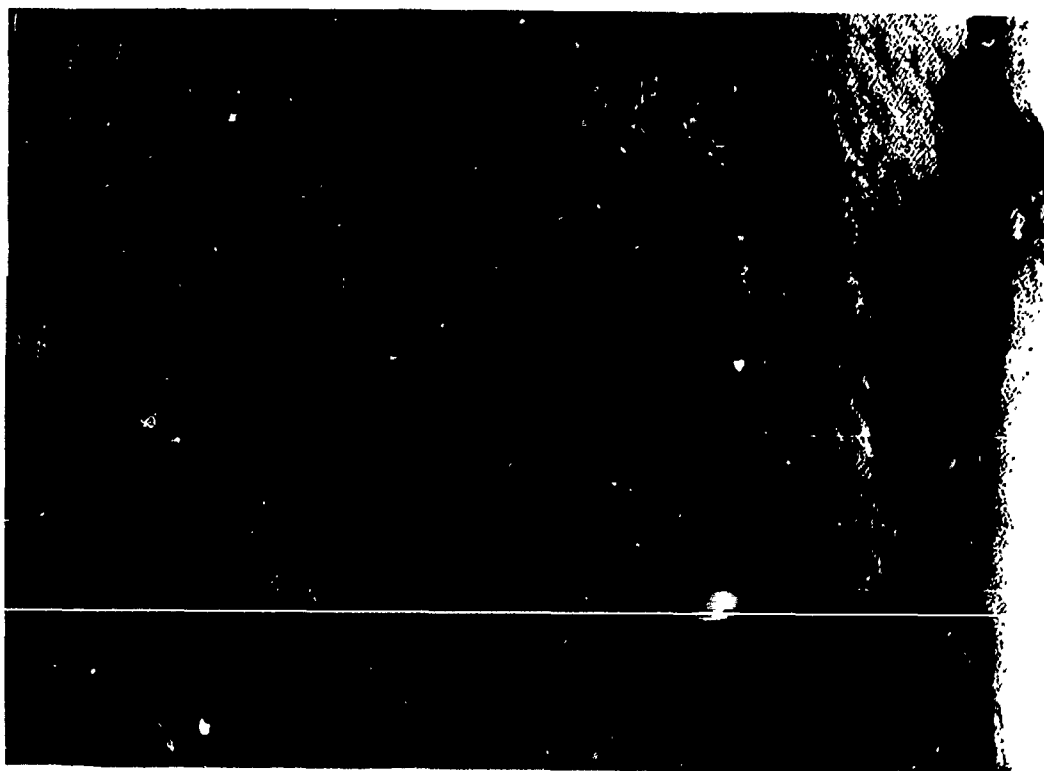
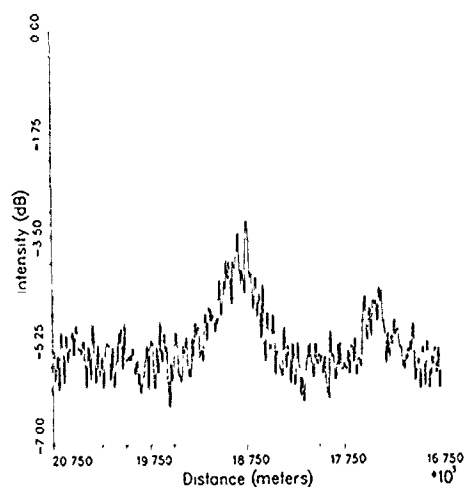


Figure 4. Enlargement of the 17 March 1988 C-Band (VV) SAR Data Showing the Current Shear Induced Ocean Front

March 13 Pass 3 C-Band Section DC



March 17 Pass 4 C-Band Section CD

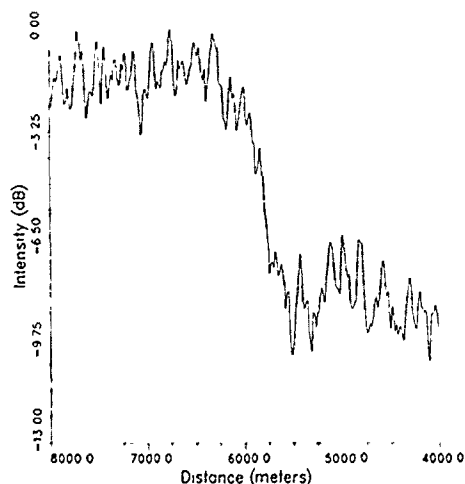


Figure 5. Intensity Plots of SAR Backscatter Along the Tracks Shown in Figures 3 and 4

3.1 MODULATION BY CURRENT SHEAR

The current vectors at 15 cm depth (integrated every 300-400 m along the ship track) obtained by the ship mounted Acoustic Doppler Current Profiler (ADCP) are superimposed on a schematic representation of the SAR image (Figure 6) documenting the present of a remarkable current shear in vicinity of the SAR front. No complementary current measurements are obtained closer to the surface. However, since the evolution of the ADCP current vectors at deeper levels down to 50 m show minor vertical current shear and rotation, the current vectors at 15 m are assumed to be representing the surface current field as well. This assumption is further supported by the thermohaline structure mapped with a towed, undulating CTD, and by the low winds measured from the ship with cup anemometers.

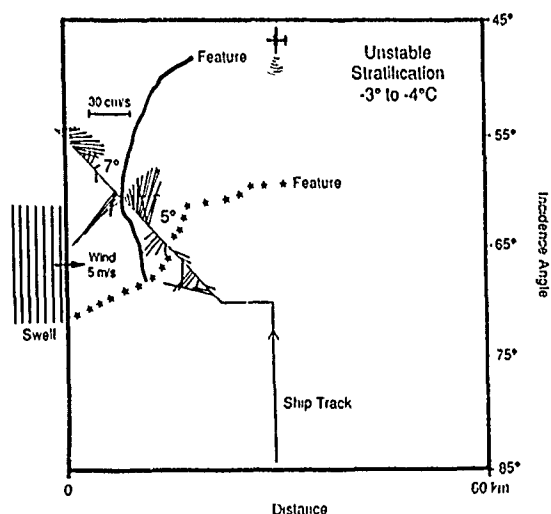


Figure 6. Schematic Representation of the SAR Image Showing the Ship Track and the Current Values Obtained from the ADCP.

In order to quantify the backscatter modulation caused by this current feature, the SAR backscatter model described by Lyzenga and Bennett [Ref. 4] is applied. This model allows us to study the effects of long surface waves (at least 3 times the radar wavelength) on the SAR imaging of ocean surface current shear zones and convergence (divergence) zones. The interaction of these surface waves with the current pattern is described by the wave action equation [Ref. 5], and solved numerically. The effects of the full spectrum of waves on the radar backscatter are then calculated by a two-scale electromagnetic surface model. Finally, the SAR image intensity and modulations are calculated using a SAR simulation model.

Model input parameters include the current vector field with a maximum current shear of 10^{-3} s^{-1} , wind direction of 45° relative to the baseline of the vector field and a wind speed of 3 m/s. The ship recorded a 5 m/s wind at the time of the SAR overflight; however, in the image the darker backscatter area in the vicinity of the front suggests that the wind speed was lower. Results of the numerical simulation were compared with the SAR observations. Only relative changes in dB are considered. The modeled backscatter modulation provides a "bump" of about 1.5 dB in favorable agreement with the direct observations of a 2 dB "bump". The effect of the relative weak current convergence (divergence) along the ship track of order 10^{-4} s^{-1} shows minor (<1dB) modulation in the model.

Increasing the wind speed to 5 m/s for the current shear case reduces the modulation by a factor of 2. Hence, the model results are extremely sensitive to the relaxation time. Nevertheless, the model results suggest that the effect of the current shear is to refract the Bragg waves originally propagating along the wind direction in azimuth direction into range propagating waves. This results in the observed backscatter increase.

Time series of the radar cross section (RCS) at 10 minutes averages along the ship track obtained by the C-band (HH polarization) ship mounted scatterometer also measured about a 1.5 dB "bump" crossing the SAR front. Simultaneously, the recorded wind speed was about 5 m/s decreasing slightly, but the direction remained constant. In addition, the boundary layer

stability remained unstable around -4°C . It is unlikely that any atmospheric effect can explain the backscatter modulation. Consequently, the SAR front is imaged as a result of interaction between Bragg waves and a current shear.

3.2 MODULATION BY A WIND FRONT

In contrast to the 2 dB SAR front modulated by a current shear, the much stronger 8 dB SAR front (Figure 4) detected on 17 March appears to be generated by a sharp wind front. In this case, the ship transected the front within 20 minutes after the feature was detected by the SAR image. The variations of the radar scattering coefficient (related to the surface roughness) are dependent on the wind speed, the atmospheric stability and long wave slope. However, under unstable conditions, Keller et al [Ref. 6] suggests that the dependence of X-band radar cross section (RCS) on long-wave slope and atmospheric stability cannot be separated from the wind speed dependence. The results discussed in the following paragraphs assume this to be valid also at C-band.

Time series of the ship based scatterometers C-band RCS and wind speed measured from R/V Hakon Mosby are presented in Ref. [7]. Two significant increments in RCS of about 7-10 dB are found with the 7 dB change in correspondence with the SAR front. In comparison, the wind speed (U) shows large fluctuations between 2 and 10 m/s. In response to these wind speed changes, the friction velocity $(U^*) = C_d^{1/2} U$ which determines the surface roughness through the relationship between the roughness scale z_0 and the drag coefficient C_d , shows a significant increment from 0.1 to 0.5 m/s. This process is reflected in the close agreement of the phase of the RCS and wind speed fluctuations.

The scatterometer data from the ship was used to quantitatively relate radar scattering coefficient to wind speed. The analysis as reported in Ref. [7] produced a relationship between RCS and wind speed under the present unstable stratification between -4.0 to -2.0°C that was supported by the results of Ref. [6]. Thus, the dB change is primarily explained by the sharp wind front. In turn, contribution to the backscatter change from the other environmental quantities must be negligible. In summary, the SAR and scatterometer measured complete backscatter modulations across the front of about 7-8 dB and are explained by the sudden wind speed change from 2 to 10 m/s.

4. SUMMARY

The analysis to date has shown that for unstable conditions and favorable low winds (4 m/s) the C-band SAR is capable of detecting current shear on the order of 10^{-3} s^{-1} that is associated with mesoscale ocean circulation features such as fronts and eddies. No SAR detection of ocean convergence (divergence) of order 10^{-4} s^{-1} along such features can be expected. However, a sharp wind front was clearly detected.

Qualitatively, the images features also appear differently according to the influence of different modulation mechanisms. SAR features imaged due to ocean circulation patterns such as a current shear frequently present along ocean fronts and eddies cause narrow backscatter changes appearing as bright lines. On the other hand, sharp wind fronts display rapid transition from dark to brighter backscatter regions. Quantitatively, one can also classify backscatter modulation caused by wind front from current shear. When the wind front exceeds 5 m/s, it can be distinguished from the current shear features simply because the step-like backscatter modulation of about 5 dB differs from the 2 dB "bump" generated by the current shear.

These results are promising for future application of ERS-1 SAR imagery to detect ocean circulation patterns independent of simultaneously obtained sea truth. Provided some background information on the strength of the ocean current and wind field are available, then mesoscale circulation features imaged by the SAR may be classified accordingly.

ACKNOWLEDGEMENTS

We thank the Canadian Centre for Remote Sensing for furnishing the digital SAR data and Catherine A. Russel for processing the images. This work was supported by the Office of Oceanographer of the Navy, SPAWAR, and the Office of Naval Research (ONR) contract #N00014-81-C-0692.

REFERENCES

- [1] Livingstone, C.E., A.L. Gray, R.K. Hawkins, and R.B. Olsen, "CCRS C/X - Airborne Synthetic Aperture Radar: An R and D Tool for the ERS-1 Time Frame", Proceedings of the 1988 IEEE National Radar Conference, IEEE 88CH2577-6, 1988.
- [2] Hasselmann, K., R.K. Raney, W.J. Plant, W. Alpers, R.A. Shuchman, D.R. Lyzenga, C.L. Rufenach, and M.J. Tucker, "Theory of Synthetic Aperture Radar Ocean Imaging: A MARSSEN View", JGR, Vol 90 C3, pp 4659-4686, 1985.
- [3] Beal, R.C., P.S. DeLeonibus, and I. Katz, Editors, "Spaceborne Synthetic Aperture Radar for Oceanography", The Johns Hopkins University Press, Baltimore, MD, 1981.
- [4] Lyzenga, D.R., and J.R. Bennett, "Full-Spectrum Modeling of Synthetic Aperture Radar Internal Wave Signatures", JGR, Vol 93 C10, pp 12,345-12354, 1988.
- [5] Phillips, O.M., "The Dynamics of the Upper Ocean, 2nd Edition", 336 pp., Cambridge University Press, New York, NY, 1977.
- [6] Keller, W.C., V. Wisman, and W. Alpers, "Tower-Based Measurements of the Ocean C-Band Radar Backscatter Cross Section" JGR, Vol 94, pp 924-930, 1989.
- [7] Onstott, R.G., R.A. Shuchman, J.A. Johannessen, O. Skagseth, and K. Davidson, "Scatterometer Measurements of Wind, Waves, and Ocean Fronts During NORCSEX", these proceedings.

SURFACE CURRENT MEASUREMENTS AT OCEAN FRONTS

Donald R. Johnson and Florence M. Fetterer

Space Oceanography Research Section
 Naval Ocean Research and Development Activity, Code 321
 Stennis Space Center, MS 39529

ABSTRACT

Using a unique slope-following surface current meter, we have obtained measurements of near surface currents during the NORCSEX experiment in March, 1988. These instruments were tethered to subsurface floatation on traditional taut wire moorings at three locations near ocean frontal areas. Several objectives of the experiment included supporting the deeper moored and acoustic doppler measurements with surface information, providing a bottom boundary condition for wind stress analysis, comparison with a CODAR and testing a newly developed ARGOS transmission capability. In this discussion, we provide a general overview of the measurement results.

KEY WORDS: SURFACE CURRENTS, SAR, FRONTS,
 WIND STRESS, NORCSEX

1. INTRODUCTION

The Synthetic Aperture Radar (SAR) has demonstrated a remarkable potential to distinguish surface features associated with ocean fronts. Our fundamental understanding of this complicated process is that features observed in SAR imagery result from wave-current interactions, which block or focus short gravity waves, and from Bragg scattering on the redistributed short wave patterns. This simplistic picture can be complicated by changes in the short gravity wave field related to wind stress changes across a front, a result of differences in boundary layer stability on each side of the thermal gradient associated with some fronts. Further complexities involve the flight direction of the SAR sensor with respect to the surface flow pattern and the bivariate slope distributions of wave facets.

Fundamental to the physics which combines hydrology and wind stress with radar backscatter, and fundamental to our ability to model these processes, is an understanding of the flow patterns at the sea surface where

strong wave/current/wind interactions occur. However, because the ocean surface is inherently noisy, it has been extremely difficult to make measurements of the ambient flow associated with open ocean fronts. Although Lagrangian drifters have supplied some information in the past, problems of deployment on scales associated with fronts as well as difficulties in separation of space and time, have clouded interpretation of the results. Acoustic Doppler Current Profilers (ADCP) are obtaining interesting observations of vertical current structure near fronts, but are not reliable in the upper 10 m of the water column due to ambiguous reflections from waves. Standard taut-wire surface moorings have provided useful information, but are difficult to maintain in the strong flows associated with fronts and may not be reliable in anything but very benign conditions.

In order to overcome some of the problems of measuring ambient flow at the sea surface, we have developed a slope-following surface current meter of a unique design which has been tested under various condition over the past several years (Johnson, 1987). During the Norwegian Continental Shelf Experiment (NORCSEX), these Rapid Boundary Current Meters (RBCM) were deployed near ocean fronts. Our objectives were:

- * to support deeper current measurements from ADCP and Aanderaa current meters with surface information,
- * to provide a bottom boundary condition for wind stress analysis,
- * to compare with surface flow measurements from Coastal Ocean Doppler Aperature Radar (CODAR),
- * to test a new ARGOS transmission capability.

2. RAPID BOUNDARY CURRENT METER

In Fig. 1, the RBCM is shown in its moored configuration. The electronics are contained within a hull which floats at the sea surface and streams with the current. In order to avoid direct wind influence on the instrument, only 2-3 cm are exposed above the surface. Orientation of the hull in the direction of ambient flow is obtained with a gymballed digital compass, and speed of the flow is measured with a near-cosine response propeller (Weller and Davis, 1980) suspended at 50 cm below the surface. An ARGOS PTT and a helical antenna allow data transfer via satellite. The instrument is moored with lightweight polypropylene line and uses small fishing floats at the surface for added buoyancy. The effect is a light, easily deployable, low stress mooring configuration. Microprocessor based electronics allows vector processing and, together with simple design, have significantly reduced the cost of the instrument and simplified maintenance.

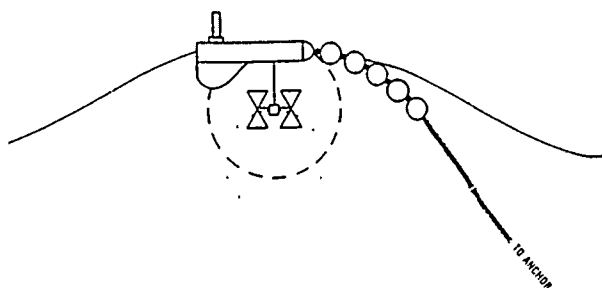


Figure 1: RBCM in its moored configuration.

The design philosophy of the RBCM has centered on a low stress, flexible mooring which permits the hull to follow the surface streamline and the propeller to follow orbital particle motions of gravity waves. In Fig. 1, wave particle orbits are drawn at the surface and at the depth of the sensor. These orbit radii decrease exponentially with depth. If the propeller followed the particle orbit at its own depth, it would tend to eliminate wave influence and to pick up Stoke's drift (Collar, et al., 1983). Since the propeller is constrained to follow the surface orbit, it is in error by a small amount proportional to the difference in size of the two orbits. However, it is clear from this simplified example, that this arrangement is a considerable improvement on a fixed level current meter in the presence of waves, or on a surface following instrument with the sensor at greater depth, or orthogonal sensors spread between two depths. The chosen depth of 50 cm represents a compromise between placing the propeller as near the surface as possible and avoiding hull induced distortion of the flow.

3. EXPERIMENT

During March, 1988, RBCM instruments were attached to the main subsurface buoys of taut-wire moorings at three locations (Fig. 2) in the NORCSEX area. Aanderaa current meters were placed at several levels on each subsurface mooring, with the most shallow instruments at 25 m depth. Wavescan buoys, recording wind and wave parameters, were also moored in the vicinity of moorings CM1 and CM2. Figure 2 shows the bathymetry in the experimental area. It should be noted that mooring CMT2 was located at the shelf break in 400 m water depth, and moorings CM1 and CM2 were located along a trough, exceeding 300 m depth, which cut across the shelf and extended behind the Haltenbanken rise.

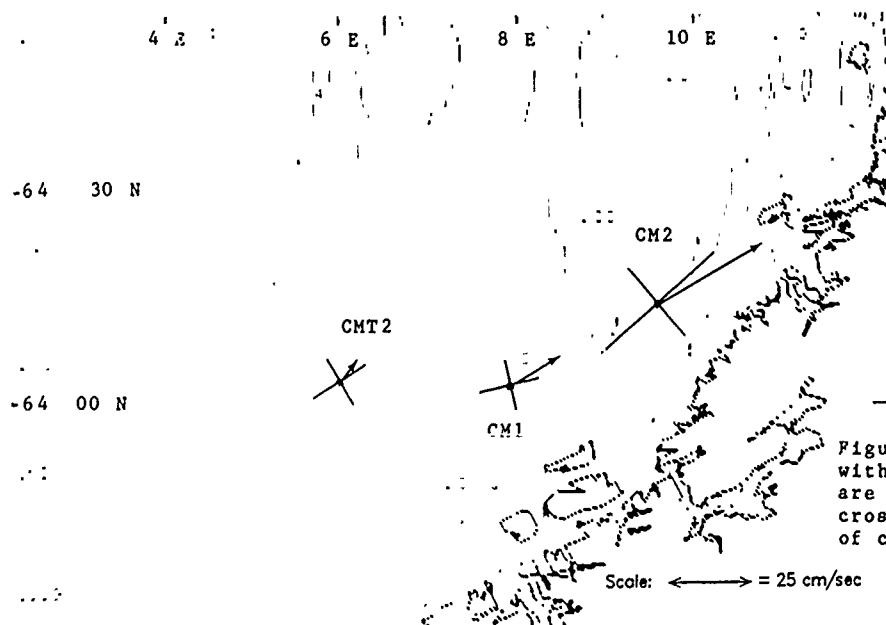


Figure 2: NORCSEX bathymetry with mooring locations. Arrows are mean current vectors; crosses are major and minor axis of current variations.

Figure 4 shows an AVHRR image for 16 March, 1988 covering the experimental area. The image has been composited from two passes, both from NOAA-9 on the 16th. Figure 5 shows an AVHRR image for 25 March, 1988 from a NOAA-10 pass. Comparing the AVHRR images with bathymetry in Fig. 2 (also lightly outlined in Fig. 4), it is clear that bathymetry played a major role in the flow regime and the separation of water masses. The Norwegian Atlantic Water, which is the warmest water in the image (darkest color), can be seen following the shelf break adjacent to CMT2. The coolest water (light color) is associated with the Norwegian Coastal Current and is found adjacent to the coastline. One of the most striking features of the two images is the curvature of the thermal boundary (front) from the shelf break, along the trough adjacent to CMT1 and CMT2, and passing behind Haltenbanken.

Vector summations of 4 hourly averages are shown in fig. 3 for the three moorings. The large gap in mooring CMT2 was caused by entanglement of the propeller. This was noted on the ARGOS transmission and corrected when ship scheduling permitted. Large subtidal variations are seen in all records. The highest surface flow speeds, averaged over a tidal period, occurred at CM2 on 20 March, with an amplitude exceeding 81 cm/s. Aanderaa records at the 25 m depth at this location (courtesy, J. Johannesssen) show a speed of 60 cm/s.

Mean velocities at each of the moorings are shown in fig. 2. In addition, major and minor axis of the subtidal flow variations are displayed with the major axis aligned in the direction of dominant variability. The mean flow vectors are all pointed toward northeast, with a near doubling of amplitude at each mooring closer to the coast. Subtidal flow variations at the shelf break are nearly omnidirectional, as seen by the fatter ellipse formed by the principal axis. As the subtidal flow becomes stronger near the coast, it is clear that the ellipses are oriented more nearly with topography.

4. SUMMARY

Three moorings of a unique slope-following surface current meter obtained a total of 36 days of current records during the NORCSEX project of March, 1988. AVHRR imagery showed that the moorings were all placed quite near the thermal boundary between North Atlantic Water and Norwegian Coastal Water. Current speeds at the surface were nearly 30% greater than speeds at the depth of 25 m. Current speeds increased dramatically near the coast and subtidal variations were aligned closely with topography. Near the shelf break, subtidal variations were surprisingly uncoupled from topography.

The easy deployment of the RBCM suggested that we should attempt to "capture" a front by chasing after the aircraft SAR image feature locations. However, deployment of a non-ARGOS transmitting instrument in this mode was unsuccessful as ship scheduling and weather combined to prevent successful retrieval. Future uses of this instrument in frontal locations should probably include more effort in this area.

Finally, the ARGOS transmission capability functioned well and, together with the relatively low instrument cost, opens the possibility of deployment of "expendable" moored current meters in the future.

REFERENCES

- Collar, P.G., R.M. Carson and G.Griffiths (1983). Measurements of near-surface current from a moored wave-slope follower. Deep-Sea Research, 30, 63-75.
- Fetterer, F.M. and D.R. Johnson (1989). A surface current meter with ARGOS telemetry capability. Marine Technology Society, Proceedings. Conference on Marine Data Systems, New Orleans, LA, April 1989.
- Johnson, D.R. (1987). A surface current meter. Continental Shelf Research, 7, 975-986.

NORDA contribution no. PR 89:035:321.
Approved for public release;
distribution is unlimited.

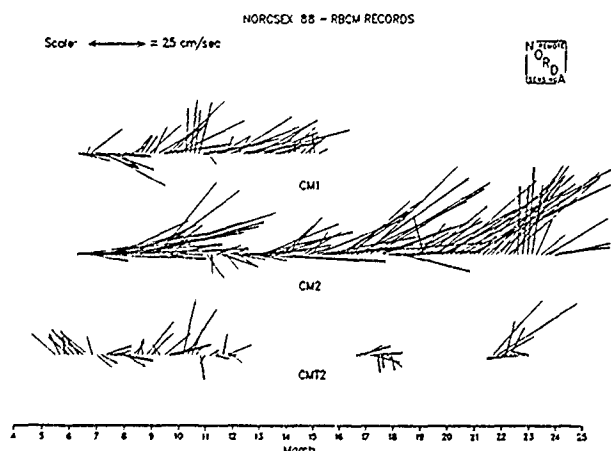


Figure 3: Stick plot of vector surface currents measured during NORCSEX. Locations given in Fig. 2.

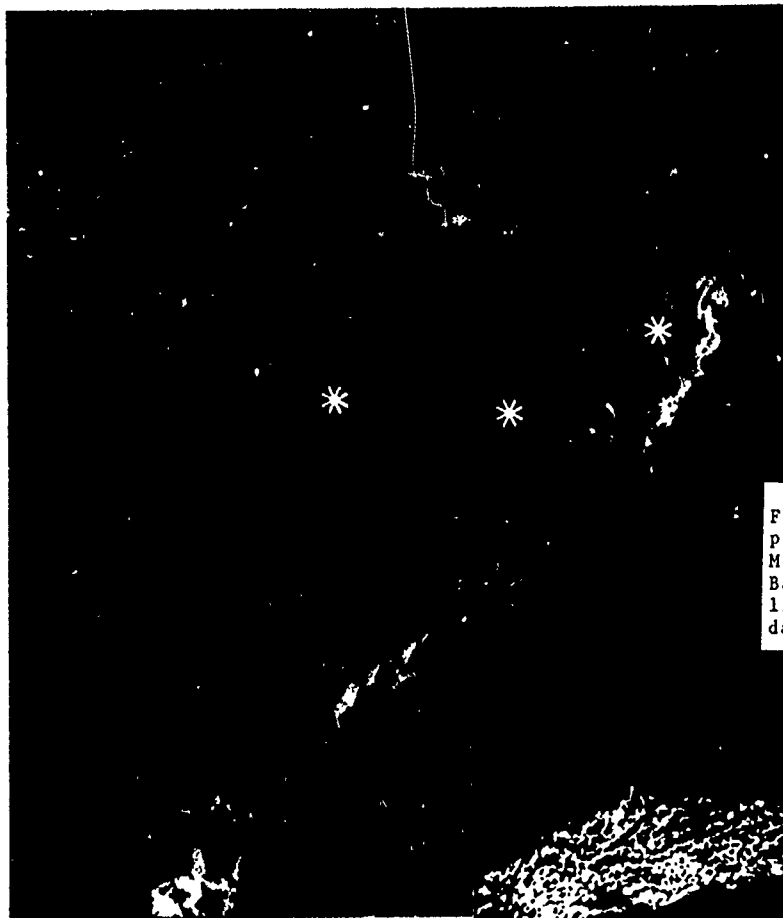


Figure 4: AVHRR composite from two passes on NOAA-9 on 16 March, 1989. Mooring locations given by asterisks. Bathymetry is overlain with dashed lines. Warm water is represented in dark tones; cool water in light.

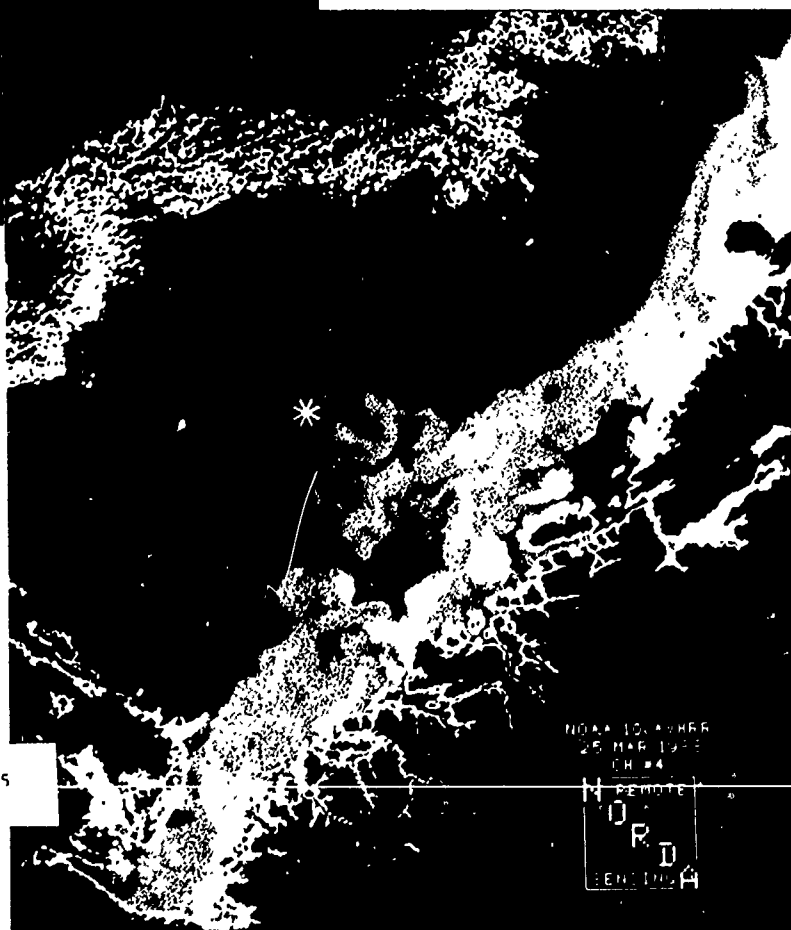


Figure 5: AVHRR from NOAA-10 on 25 March, 1989.

SURFACE CURRENTS DURING NORCSEX'88, AS MEASURED BY A LAND- AND A SHIP-BASED HF-RADAR

H.-H. Essen, K.-W. Gurgel, F. Schirmer, T. Schlick

Institut für Meereskunde, Universität Hamburg
Tropelwitzstr. 7
D-2000 Hamburg 54 (Germany)

ABSTRACT

CODAR (COastal raDAR) measurements of surface currents, carried out during NORCSEX'88, are presented. One land- and one ship-based station were used, in order to map the two-dimensional current off the Norwegian coast at about 64°N and 8°E in March 1988.

The ship-based CODAR system, developed by the University of Hamburg, is new and discussed in some detail. Errors due to angular resolution by means of an array of four receiving antennas, determination of the ship's drift by satellite navigation GPS (Global Position System) and pitch-and-roll motions of the ship sum up to about 10 cm/s. This value is corroborated by the comparison of current velocities measured by the land- and ship-based system at the same position.

Surface currents observed show a high temporal and spatial variability on scales of several hours and some km, respectively. Temporal variations are obviously due to wind. The horizontal variability of the Norwegian Coastal Current is well known from satellite images. The underlying mechanisms are complex and not yet understood.

1. INTRODUCTION

CODAR is an HF ground-wave radar, which is able to measure surface currents in the upper 0.5 m of the ocean. An area of up to 40 km x 40 km may be covered, with a horizontal re-

solution of about 2 km. The time averaging is 18 minutes (measuring time). Originally, CODAR was developed by NOAA for land-based operation (Barrick et al., 1977). The system has been operated successfully in a number of experiments (Gurgel et al., 1986; Schott et al., 1986).

The use of CODAR allows to measure mesoscale surface currents fields, which are hard to obtain by conventional means. Airborne remote sensing systems (optical, infrared, microwave) show sea-surface structures, which are obviously due to currents. CODAR offers a unique method of supplying these systems with ground-truth data, which are necessary for understanding the imaging mechanisms and for developing processing algorithms.

A ship-based system allows to operate in open water, which is of great interest for the research at our institute. First experiments in 1984 were encouraging but lacked through the unknown velocity of the ship. This problem was overcome, when GPS data became available for civil use.

2. SHIP-BASED CODAR

Before introducing the ship's CODAR, we shortly present the basic ideas behind the system.

The CODAR system transmits pulses, allowing resolution in range. The signal is partly backscattered by ocean surface waves with one half of the radar wavelength, running

towards or away from the radar site (Bragg scattering).

The Doppler-shift of the backscattered signal yields the phase velocity of the scattering surface waves, which is composed of the theoretically known phase velocity in nonmoving water and the respective component of the underlying current velocity. Two radar sites at distant positions, each measuring radial velocities, are necessary to determine two-dimensional current vectors.

The CODAR-system used works with an (nearly) omnidirectional transmit antenna. The azimuthal resolution is performed by means of an array of four receiving antennas, arranged in a square. In principle, two incident angles may be resolved for each frequency of the Fourier spectrum.

With operating the CODAR from board a ship, a number of problems arise:

a) The ship's metal perturbs the electromagnetic field, and the angular resolution cannot be carried out by means of the free-field solutions. This problem is overcome by using measured antenna characteristics, from which we only regard the phase differences between the antennas. It turned out that these values are similar to the free-field ones and allow a unique azimuthal resolution.

From phase differences, only one incident angle may be resolved. But within a circle, the same radial velocity occurs from at least two directions. In order to avoid this ambiguity, we transmit into a semicircle and move the ship with low velocity (about 1 kn).

b) The ship is drifting, and the measured current velocity has to be corrected with the ship's velocity. For this purpose, a GPS is integrated into the CODAR. Its accuracy mainly determines the accuracy of the obtained current velocities.

In case of optimal satellite coverage, GPS yields ship velocities of high accuracy with rms-values below 5 cm/s. This estimate is obtained for a fixed position (zero velocity).

Because of pitch-and-roll motions, the measurements on board the ship yield higher values. On board R/V KRONSORT, the ship operating CODAR during NORSCSEX'88, the GPS-antenna was installed on top a mast 20 m above sea surface, for which roll motions caused amplitudes up to 3 m. These were registered by GPS and a comparison with accelerometer data showed excellent agreement. Otherwise, pitch motions with amplitudes below 1 m, could not be resolved by GPS.

Our measurements showed that the ship's drift cannot be kept constant during the measuring time of 18 minutes, but varies slowly with amplitudes of 10 cm/s or even more. For considering this effect, the time series are divided into 8 and additional 7 overlapping parts, for which radial current velocities are computed and corrected by the actual ship's drift.

c) The pitch-and-roll motions of the ship induce amplitude and phase modulation of the received signal. Calculations with synthetic data show that the respective Doppler lines become relevant in case of antenna movements of the order of the electromagnetic wavelength (10 m). Transmit and receive antennas were installed at about half the height of the GPS antenna, that means they moved up to 1.5 m in both directions. As theoretical results show, this is acceptable to the assumed accuracy.

In this context, another modulation, also affecting the land-based system, should be mentioned. The relatively short scattering surface waves (5 m) may be carried by long waves (swell) causing additional Doppler-shifts due to their orbital motion. To a certain extent, this effect is smoothed by averaging over the number of waves covered by a radar pulse (2.4 km).

Fig. 1 shows radial current velocities as obtained from the two radar sites. The line marks the connection of the land- and ship-station, where both should measure the same velocity. Allowing deviations of the order of 10 cm/s, this was the case throughout the experiment. The example of Fig. 1 shows opti-

mal ranges of both stations, 40 km for the ship radar and 50 km for the land radar. Depending on sea condition and radio noise, the ranges decreased to 30 km and 40 km, respectively.

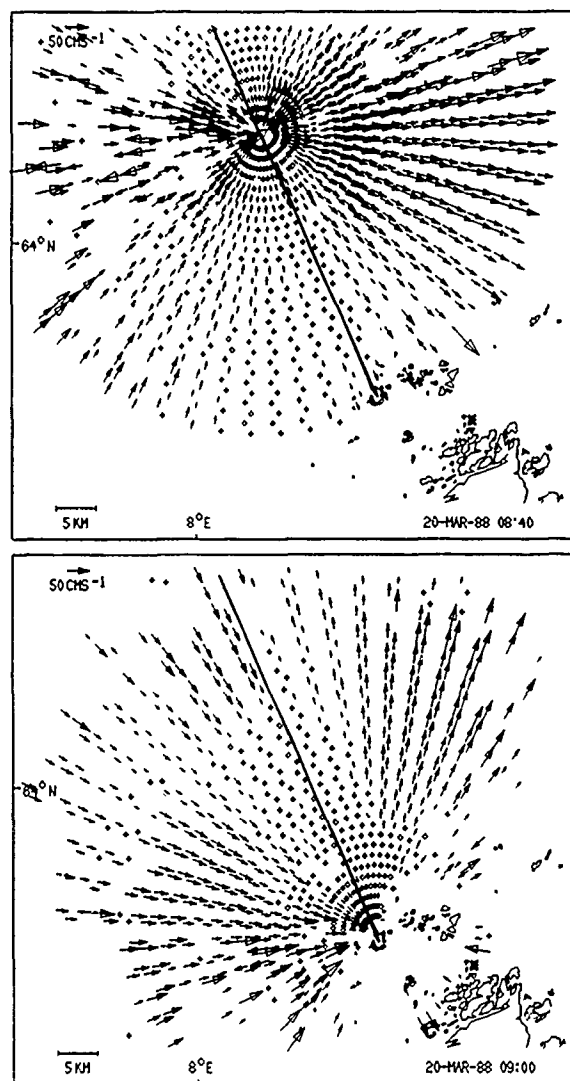


Fig. 1: Radial current velocities, measured by CODAR from board R/V KRONSORT (upper panel) and the island SULA (lower panel). In order to avoid interference, the measurements were carried out successively. The straight line connects both sites.

3. SURFACE CURRENTS DURING NORCSEX'88

CODAR data during NORCSEX'88 were obtained from two sites, one land-based on the island SULA, the other ship-based on R/V KRONSORT. The experiment was performed from 14 March, 7:00 to 22 March, 11:00 NLT, 1988.

While the land-based system worked every hour, the ship site could be operated only for five hours in the morning and six hours during evening, when GPS data were available. In order to avoid interference, the measurements from both sites were carried out successively around full hour. A few measurements are lacking because of system failure. Due to bad weather conditions, no reliable data were obtained on 18 March, morning.

R/V KRONSORT was operated from two different positions, about 30 km west of SULA for the first days of the experiment and afterwards the same distance north of SULA.

Fig. 2 shows one two-dimensional current map from each position. The grid distance is 3 km. Each vector is obtained from radial velocities measured within a circle of 3 km radius around a grid point. That means, velocities of adjacent grid points depend on each other. Around the connecting line of both stations, no two-dimensional vector may be constructed, because both stations measure the same component.

The first example in Fig. 2 (upper panel) was chosen, because a satellite SST (Sea Surface Temperature) image is available for that time. There is some obvious similarity with the CODAR map. The frontal structure of surface currents coincides with gradients of 1°C in SST.

The second example in Fig. 2 (lower panel) shows two-dimensional currents computed from the radial velocities of Fig. 1. This map shows the general trend of current flowing north-east parallel to the shelf.

Maximum current speeds are of the order of 50 cm/s. This agrees with the somewhat smaller velocities observed below surface by means of moored Aanderaa current meters and

a ship-borne ADCP (Acoustic Doppler Current Profiler).

Unfortunately, the RBCM (Rapid Boundary Current Meter) within the CODAR area failed on 15 March. The remaining data for comparison

from 14 March are low velocities at the sensitivity limit of the RBCM. Nevertheless, the agreement is good.

As CODAR maps show, surface currents may considerably change speed and direction on scales of some 10 km. Structures, like meanders, fronts and eddies are visible. These features of the Norwegian Coastal Current have been observed in satellite images since several years. But, contrary to CODAR, these images contain no information on current velocities.

From one measurement to the next, that means on hourly scales, there are only slight changes in the current field. But, in accordance with the wind field, currents may vary considerably within several hours. During the experiment, wind speeds ranged from 0 to 12 m/s with different directions. Periods of relatively stable winds took between 6 and 24 h.

4. CONCLUSIONS

During NORCSEX'88 the ship-based CODAR worked successfully with an accuracy of 10 cm/s. The currents fields observed show high meso-scale variability, also present in satellite images. The interpretation is still on the way. Especially, a comparison of CODAR with SAR (Synthetic Aperture Radar) data is lacking.

REFERENCES

1. Barrick, D.E., Evans, M.W., Weber, B.L., "Ocean surface currents mapped by radar", *Science*, Vol. 198, pp 138-144, 1977.
2. Gurgel, K.-W., Essen, H.-H., Schirmer, F., "CODAR in Germany - a status report valid November 1985", *IEEE J. Oceanic Eng.*, Vol. OE-11, pp 251-257, 1986.
3. Schott, F.A., Frisch, S.A., Larsen, J.C., "Comparison of surface currents measured by HF Doppler radar in the western Florida Straits during November 1983 to January 1984 and Florida Current transports", *J. Geophys. Res.*, Vol. 91(C7), pp 8451-8460, 1986.

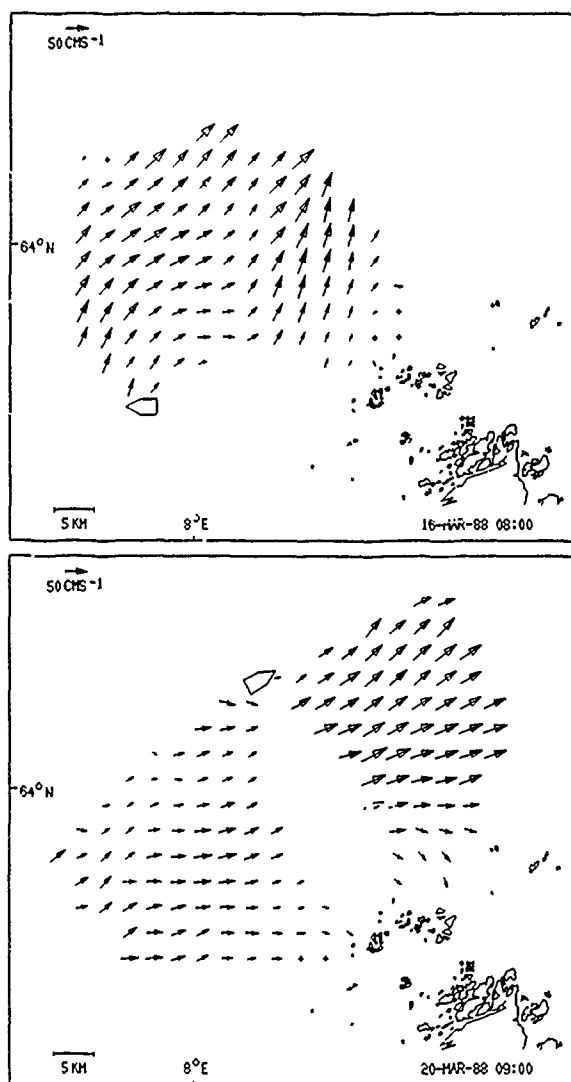


Fig. 2: Maps of two-dimensional surface currents measured by CODAR. Shown are examples from both positions of R/V KRONSORT. For the upper map a satellite SST image is available, the lower map is constructed from the radial velocities of Fig. 1.

MESOSCALE VARIABILITY IN THE HALTENBANKEN REGION MAPPED BY GEOSAT TOPOGRAPHY DATA

L.H. Pettersson, J.A. Johannessen, T.A. Johansen and P. Samuel

Nansen Remote Sensing Center
Edvard Griegsvei 3A,
5037 Solheimsvik - Bergen,
Norway

ABSTRACT

Repeated track analysis of the GEOSAT altimeter height data from the entire Exact Repetition Mission (ERM) have been used to reconstruct the mesoscale variability of the sea surface topography in Haltenbanken region, off the west coast of Norway.

The unique data base, obtained during NORCSEX'88, allow us to investigate the capability of the GEOSAT altimeter to resolve the mesoscale ocean height variations (approximately 20 cm), in particular corresponding to the existence of meanders and eddies.

Current measurements by a shipmounted Acoustic Doppler Current Profiler (ADCP) and simultaneous NOAA AVHRR images provide quantitative estimates of horizontal extension and orbital speed of these mesoscale circulation features. The analysis of these data suggest that the variability in the ocean currents and corresponding mesoscale height variations can be resolved by the GEOSAT altimeter height data.

Key words: ALTIMETER, GEOSAT, OCEAN CIRCULATION, EDDIES, ADCP, AVHRR

INTRODUCTION

Haltenbanken is located on the continental shelf off the west coast of central Norway. The mesoscale ocean circulation in this region is affected by a mixture of several mechanisms, including topographic steering by the shelf break and depressions and seamounts on the wide continental shelf as well as instability at the boundary between the Norwegian Coastal Current and the North Atlantic Current. In consequence generation of mesoscale meanders and eddy features occur regularly.

OBSERVATIONS

The GEOSAT altimeter investigation during NORCSEX'88 is part of a preparation study for the ERS-1 and Topex/Poseidon radar altimeter missions.

The sea surface topography data from the US GEOSAT altimeter satellite have been analyzed by repeated track analysis for the entire Exact Repeating Mission (ERM).

The GEOSAT satellite completed almost two 17 days repeat cycles during the duration of the NORCSEX'88 field campaign in March 1988. For this particular period an extensive set of oceanographic and meteorological data were simultaneously obtained from research vessels, moorings, drifting buoys, aircrafts, and NOAA satellites (Johannessen et al., this issue). Comparison of the altimeter resolved variations in ocean surface topography with this extensive set of *in situ* and remote sensing data allow us to evaluate the altimeter measurements.

A series of cloud free images from the NOAA satellites resolved the horizontal scales of the eddies and meanders in the range of 50 to 100 km. The Acoustic Doppler Current Profiler (ADCP) on-board the R/V Haakon Mosby sampled continuously the absolute current from the surface layer (10 m below surface) to the bottom, which varied from a depth of 100 m to 500 m. These current measurements are analyzed by objective methods (Haugan et al., this issue) in order to construct the synoptic mesoscale variability in the ocean currents. Combination with current data from fixed moorings allow separation of the tidal current component from the ADCP data. The horizontal scales, obtained by the ADCP are in agreement with the typical meander size resolved in the AVHRR images. In addition the change in speed is 50 cm s^{-1} across these features of about 30 km. The corresponding change in sea surface topography is

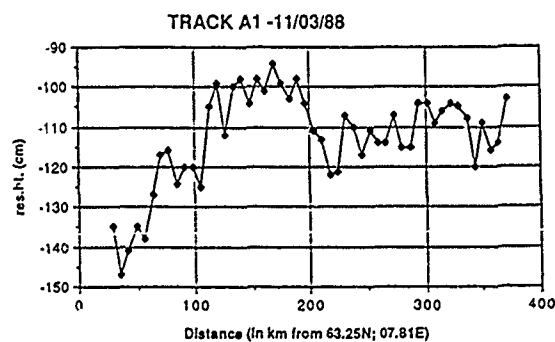


Figure 1: The residual height of the ocean topography as resolved by repeated track analysis of GEOSAT altimeter height data.

about 20 cm when the geostrophic balance equation is applied. Since the vertical velocity structure is observed to have a significant barotropic component the observations of sea surface slopes (heights) and subsequently comparisons to the altimeter ocean topography measurements are valid. In Figure 1 the residual height of the altimeter measured topography for an ascending track is shown. A significant anomaly 20 cm in height and 100 km in width are seen.

In order to construct a horizontal contour map of the residual height, 8 ascending and 6 descending tracks were selected. Data from the first 36 repeat periods of the GEOSAT ERM were used in the analysis. Each track is typically a few hundred kilometers long within the box considered but data gaps occur, leading to variation in the extent of the track from one period to the next. Small lateral deviations are also present in the repeat tracks. To facilitate averaging, equidistant points were defined on a standard track and trend-removed height data were interpolated to these points from the closest data points from each repeat period. Then, at each point the mean height and the standard deviation were computed. For the two NORCSEX periods, the mean heights were subtracted from the original trend-removed data to give the mean deviation.

After the data from each track were processed in this manner, a plotting program was used to map the standard deviation (Figure 2), which give indications of the mesoscale variability in the region.

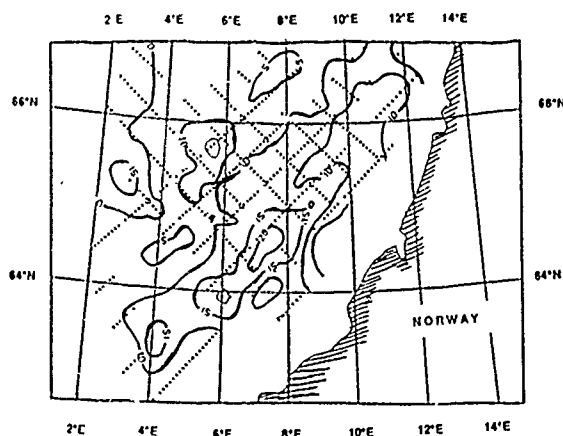


Figure 2: Standard deviation of trend removed altimeter height.

It is possible that the trend-removal operation may have led to the loss of part of the mesoscale signature, considering the short length on a few of the tracks. It is proposed that longer tracks should be used in future analyses. Also no allowance has been made for the eventuality that there might exist significant differences in the data around the cross-over points from the ascending and the descending tracks.

The contour plot is elongated in the descending direction almost parallel to the coast. This suggests

that more variability occur along the ascending track direction. This is expected since the mean flow direction of the current system is north eastward. The variability range from 5 to 20 cm. Maximum standard deviation is located where the Norwegian Coastal Current accelerates north-eastward due to a "bottle neck" effect induced by a seamount (Haltanbanken) offshore.

Finally, a full analysis using all the tracks from the area would yield a more comprehensive picture of the mesoscale variability obtainable from satellite radar altimetry.

MESOSCALE VARIABILITY DURING NORCSEX '88 DERIVED FROM AVHRR SURFACE STRUCTURES, ADCP CURRENTS AND A NUMERICAL MODEL

P.M. Haugan*), J.A. Johannessen*), M. Ikeda**) and K. Kloster*)

*)Nansen Remote Sensing Center
Edvard Griegsvei 3A, 5037 Solheimsvik
Bergen, Norw .

**) Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, N.S. B2Y 4A2, Canada

Combination of remote sensing and in situ data sets with each other and with model results are used for description and understanding of the mesoscale variability in the area around Haltenbanken off mid-Norway. This is an eddy-rich region, where the generally northward-flowing Atlantic water meets a widening continental shelf and mixes with coastal water.

Data collected during NORCSEX in March 1988 allows us to demonstrate the correlation between sea surface structures as measured by NOAA AVHRR, and in situ measured currents along the meandering front between warm Atlantic water and cold coastal water. The AVHRR imagery is supplied with grid information, screened for clouds, and then processed to enhance the temperature gradients. In situ current measurements are obtained from shipborne Acoustic Doppler Current Profiler (ADCP), providing large areal coverage and high vertical resolution in a short survey time. Bottom-tracking of the ADCP on the shelf allows reliable determination of absolute currents. Synoptic maps are obtained through objective analysis of the ADCP data, utilizing also time series from moored current meters to model the tidal components of the current variability.

Following an initial comparison between AVHRR sea surface temperature structures and in situ currents, the AVHRR is subsequently used as additional data to improve the objective analysis scheme for current mapping from ADCP. The analysis of the current field is constrained by requiring the preservation of fronts found in AVHRR images.

A three-layer quasi-geostrophic numerical model covering 300km x 200km of the experiment area, is set up to study processes responsible for meandering and eddy generation along the shelf break and in the Norwegian Coastal Current.

The model has idealized bottom topography including the shelf break, the northward widening and deepening of the shelf, and the Haltenbank represented as a sea-mount. Instabilities in the basic configuration and interactions with topography are explored, and compared with the observed flow patterns. Numerical model simulation will be attempted based on the model results from the idealized current structures and bottom topography.

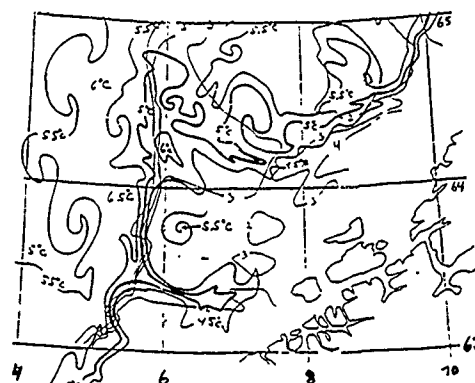


Fig. 1 Surface temperature from AVHRR on 25 March



Fig. 2. ADCP currents at 25m on 25 - 26 March

ANALYSIS OF GEOSAT ALTIMETER SEA STATE OBSERVATIONS AND COMPARISON WITH SIMULTANEOUS AIRCRAFT SAR OBSERVATIONS FOR NORCSEX-88

Nelly M. Mognard
Centre National d'Etudes Spatiales
18 Avenue E. Belin
31055 Toulouse-Cedex France

Charles Livingstone
CCRS
2464 Sheffield Rd.
Ottawa, Ontario K1A 0Y7 Canada

Johnny A. Johannesen
Paul Samuel
NRSC
Edvard Griegsvei 3A
5037 Solheimsvik/Bergen, Norway

Robert Shuchman
ERIM
P.O. Box 8618
Ann Arbor, MI 48107 USA

ABSTRACT

During NORCSEX-88 (NORwegian Continental Shelf EXperiment) in the Haltenbanken region off the coast of Norway in March 1988, a quasi-simultaneous aircraft underflight of the Geosat altimeter took place with the X- and C-band SAR on board the Canadian Convair 580. This flight transected a well-developed storm in which high waves and winds occurred.

Analysis of the Geosat altimeter sea state parameters are presented and compared with SAR and wavescan buoy data. The Geosat sea surface wind speed and significant wave height measurements are used to compute minimum significant swell heights, and parametric JONSWAP wave spectra. A swell family, sequentially observed by the Geosat altimeter, propagated from a region southwest of Scotland to the Haltenbanken region, where it was observed simultaneously by the altimeter and the aircraft SAR. Swell wavelength deduced from the Geosat altimeter, the airborne SAR, and the wave rider buoys are compared.

1. INTRODUCTION

On March 20, 1988, as part of the NORCSEX-88 experiment, the Canadian Center for Remote Sensing (CCRS) Convair-580 (CV-580) Synthetic Aperture Radar (SAR) system flew a data acquisition pattern over the Haltenbanken region of the Norwegian continental shelf to provide supporting data for ocean surface measurements made during an ascending pass of the Geosat altimeter. Figure 1 identifies the major features of the aircraft flight pattern and the location of the four directional wave rider and meteorological buoys that monitored sea state and atmospheric conditions during the whole NORCSEX-88 experiment. The SAR flight tracks on this diagram are 34 degrees incidence angle tracks with a 10 nautical miles (n mi) wide imaged swath which starts

at the aircraft nadir. The first line (1/1) was synchronized with the Geosat overflight time so that the two data sets would be coincident in time and space in the vicinity of buoy 4. The second line (2/2) paralleled the first with the radar look direction reversed in space to provide confirmation of line 1 data and to monitor temporal changes in the ocean surface. Flight lines 3 and 4 were flown to provide multiple aspect angle data in the vicinity of buoy 4 and to link data from buoy 1. During all flight lines, the X- and C-band SAR's were operated in VV polarization using the nadir mode of the system to provide surface measurements over a large incidence angle range (0 to 74 degrees).

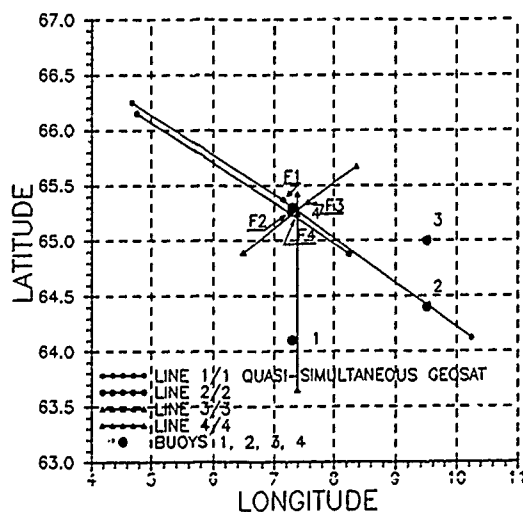


Figure 1: CCRS SAR CV-580 flight lines pattern on March 20, 1988 in the NORCSEX-88 area, with the locations of the frontal feature SAR observations along each flight line (F1 to F4), and the buoys locations.

The simultaneous airborne SAR and Geosat altimeter data were acquired during a local storm that started on March 19 with strong winds blowing from the southwest along the coast. During the Geosat pass at 8 hours UTC on March 20, an occluded front was located over the Haltenbanken region, parallel to the Norwegian coast.

The SAR imagery obtained on March 20 is characterized by well-defined wave fields throughout all passes and systematic slow variations in radar return with sudden discontinuities at the crossing of the frontal feature in the vicinity of buoy 4 (marked F1, F2, F3, and F4 in Figure 1). This front is very sharply defined and is seen by the SAR at all incidence angles. This SAR feature is associated with an atmospheric front and not with a front or a current boundary in the ocean surface layer. Similar features observed during other NORCSEX flights were probed with oceanographic sensors by the research vessel *Hakon Mosby* and were found to be sea surface manifestations of surface wind structures (pronounced wind shear). During the time interval between the front crossings in data acquisition passes 1 and 2 (62 minutes), the front had shifted 16 n mi from position F1 to F2 (Figure 1). The shift was verified in flight line 3 and line 4 (positions F3 and F4 in Figure 1).

The dominant swell field observed in the SAR imagery came from the southwest and was augmented by a wind sea especially to the east of buoy 4.

2. THE CCRS CV-580 SAR SYSTEM DURING NORCSEX-88

The CCRS CV-580 C- (5.30 GHz) and X-band (9.25 GHz) SAR system was the primary data acquisition sensor during the NORCSEX-88 flight program from March 11 to March 21, 1988. During this period, the C-band SAR was nearing the end of its commissioning phase but the X-band was newly installed and provided its first extensive ocean imagery. The radar systems are described in detail in Livingstone et al. (1987) and in Livingstone et al. (1988).

Both C- and X-band radars employ selectable transmitter polarization and simultaneous phase-coherent, dual-polarized receivers to allow the acquisition of simultaneous and cross polarized images. Both antennas are carried on the same two axis (azimuth/elevation) controlled drive, they view the same terrain and cannot be pointed independently. For the NORCSEX experiment, both radars were operated exclusively in a VV polarized single channel mode for ocean measurements. The wide swath imaging mode (18 m range resolution, 61 km swath at 15 m range pixel separation) was used for investigation of large scale sea surface structure and the nadir mode (5.7 m range resolution, 16.4 km swath at 4 m pixel separation) was used to measure ocean waves. The nadir mode geometry allows imaging from incidence 0 to 74 degrees and the real-time processed image is normally recorded in the slant range plane to minimize field errors arising from errors in aircraft attitude estimation.

3. THE GEOSAT ALTIMETER SEA STATE MEASUREMENTS

The Geosat altimeter is a 13.5 GHz nadir-looking pulse compression radar that performs three independent measurements:

- 1) the altitude of the altimeter above the ocean surface derived from the time delay between transmission and reception of the radar altimeter signal;
- 2) the significant wave height (H1/3) computed onboard the satellite from the slope of the return waveform leading edge; and
- 3) the surface wind speed (U) derived from the Automatic Gain Control loop used to normalize the amplitude of the ocean return signal.

The Geosat radar altimeter system is described in MacArthur et al. (1987). The H1/3 and U measurements are acquired along the satellite track every second (over a distance of approximately 7 km) and are used for sea state analysis. The precision of the H1/3 measurement is 50 cm rms or 10 percent of H1/3, whichever is greater, and the wind speed precision is 1.8 m/s rms up to 18 m/s (Dobson et al., 1987). The algorithm selected to compute the wind speed is described in Brown et al. (1981).

From the Geosat H1/3 and U measurements, a minimum significant swell height can be derived when the H1/3 measured by the altimeter is greater than the fully developed wave height derived from U. The fully developed wave height is computed using the formulation in Pierson and Moskowitz (1962). When the fully developed wave height associated with U is smaller than the altimeter-measured H1/3, the difference between the energy in the altimeter H1/3 field and the fully developed wave field is due to the presence of swell, and a minimum significant swell height can be computed (Mognard, 1984).

4. SAR AND ALTIMETER OCEAN SCATTERING CROSS SECTION

4.1 THE SAR MEASUREMENTS

An analysis of the along-track variation in radar reflectivity at constant incidence angle is performed using outputs from the C-band SAR real-time processor. The data are block averaged over 1.2 km ground range by 6 km along track interval centered at 34 degrees incidence angle (Figure 2a). The real-time output of the SAR has an image signal magnitude that is a monotonic function of the mean ocean scattering cross section and thus can simply be related to surface wind speed.

Gray and Hawkins (1985) show that at C-band the relationship between radar cross section (σ) and wind speed (U):

$$\sigma = CU^7$$

is valid over a wide range of wind speeds and incidence angles if C is a function of incidence angle and τ a function of both wind speed and incidence angle.

At off-nadir angles, Bragg scattering is the dominant physical effect that determines the return signal to the radar imaging the ocean surface. An increase in radar reflectivity corresponds to an increase in wind speed, such as the case of the strong wind-driven sea in the vicinity of buoy 2 near the Norwegian coast (Figure 2a), and a decrease in radar reflectivity to a decrease in wind speed, as in the case of the sharp decrease of 7 dB in radar reflectivity at the crossing of the frontal feature F1 (Figure 2a).

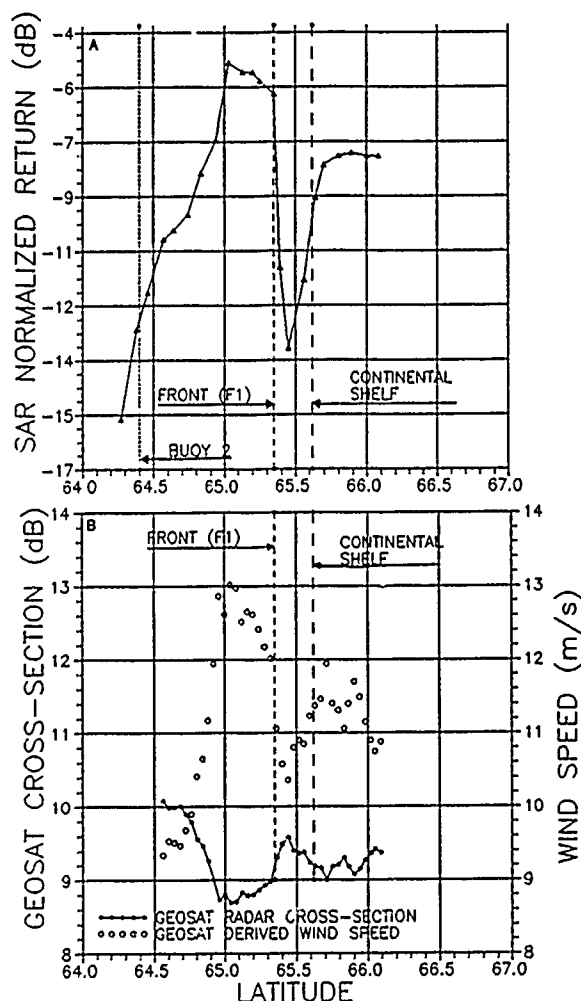


Figure 2: (a) Variations of the SAR normalized return along flight line 1, location of buoy 2, of the frontal feature (F1) and of the edge of the continental shelf; (b) quasi-simultaneous variations of the Geosat altimeter cross-section and of the derived wind speed.

4.2 THE GEOSAT ALTIMETER MEASUREMENTS

For the Geosat altimeter, the amplitude of the ocean-return signal is normalized via an Automatic Gain Control (AGC) loop. The AGC setting measures the backscatter coefficient at the ocean surface which is dependent on wind speed (Mognard and Lago, 1979). The scattering cross-section of the ocean surface at nadir can be modelled by specular points assuming that the sea surface slopes are nearly gaussian and isotropic in their distribution (Barrick, 1974). The altimeter senses the variations in specular reflection along the track and thus responds to changes in wind speed that are opposite to the SAR responses.

During NORCSEX-88 on March 20, the variations of the Geosat altimeter ocean cross-section along with the derived wind speeds are shown in Figure 2b. The position of the front (F1) is sharply defined by an increase of 0.6 dB on the altimeter cross-section (Figure 2b).

The Geosat pass starts north of buoy 2, which has a sheltered position near the Norwegian coast (Figure 1). On March 20 at 5 hours UTC, buoy 2 measured a wind speed of 8.5 m/s, and at 8 hours UTC a speed of 8.9 m/s, both with a stable wind direction from the south-southwest. A local fetch-limited sea was present at buoy 2 where the wind speed measurements are in good agreement with the Geosat wind speed at 9.3 m/s at the start of the pass. Buoy 4, which is located at the edge of the atmospheric front at the time of the Geosat pass, recorded very variable wind speeds of 20.5 m/s at 6 hours UTC and 5.3 m/s at 9 hours UTC. Buoy 4 was measuring very variable wind conditions corresponding to an atmospheric front crossing.

5. SAR AND ALTIMETER WAVE MEASUREMENTS

5.1 THE SAR PROCESSING

SAR images of ocean waves are sensitive to SAR system parameters, such as incidence angle, radar frequency, antenna look direction relative to the wave field, integration time, and slant range distance over motion of the SAR platform (R/V ratio), as well as wind speed, long wavelength wave field, wave height and slope, and surface currents.

A SAR preprocessing is performed onboard the aircraft. This processing uses the onboard navigation system to maintain good spatial control of the image locations. The radar system real-time motion compensation is engaged for all the flight lines. The real azimuth correlator provides a focussed image output in flight. This unit performs a time domain correlation between seven frequency defined sub-beams and a stored azimuth reference function (computed for the radar imaging geometry and initial ground speed at the start of each flight line) then recombines the time shifted, detected look images to form an output magnitude image.

A complex precision processing is performed on the ground. This processing consists of:

- 1) slant to ground range conversion, (geometric correction necessary because the SAR views the ground at an oblique angle);
- 2) windowing and median filtering of the data for speckle reduction;
- 3) two-dimensional Fast Fourier Transforms (FFT) assuming that the ocean surface can be represented as a stochastic field; and
- 4) application of a modulation transfer function to the data (Monaldo and Lyzenga, 1986).

Plots of the wave number-intensity spectra provide information about wavelength and wave travelling direction along the track. Ground corrected area with 512x512 pixel size have been selected in order to approximately contain 10 waves of 200 m long wavelength. Wave numbers in azimuth and range are indicated along the axis, while the circles identifies the wavelength for every 100 m between 100 and 500 m. In order to obtain the wave travelling direction in earth coordinates, the axis must be rotated clockwise so azimuth becomes parallel to the flight direction.

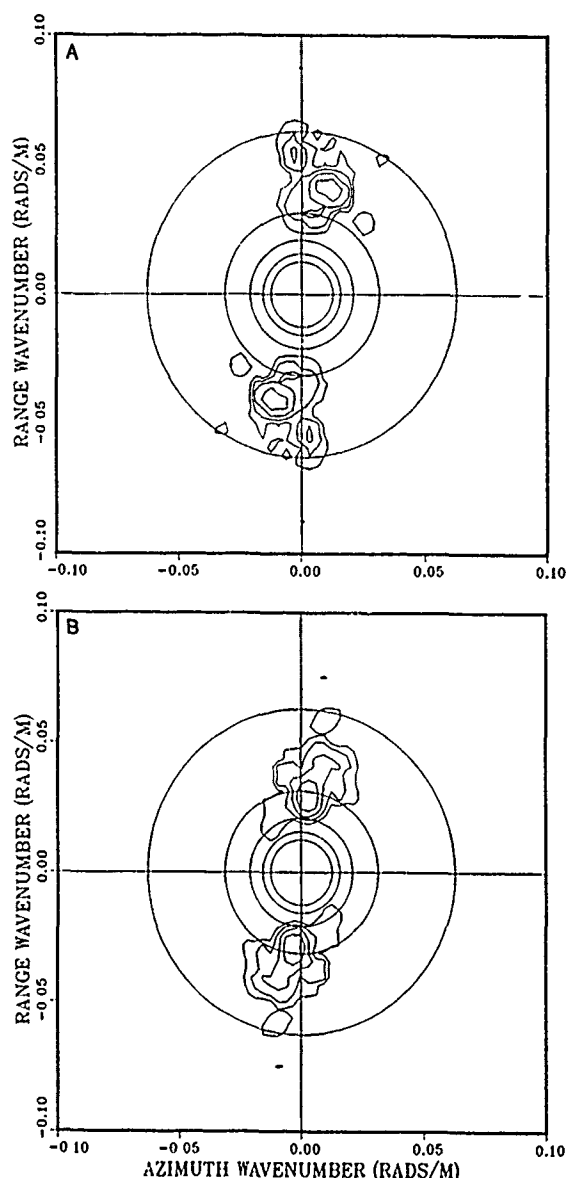


Figure 3: C-band SAR wave number-intensity spectra along flight line 1 at a) latitude 64°52', longitude 8°33', and at b) latitude 66°03', longitude 5°26'.

The radar look direction was favorable for flight lines 1 and 2 on March 20, imaging mostly range travelling waves. These waves had a mean travelling direction of 230 degrees with a mean wavelength of about 180 m. Figure 3a and 3b gives two examples of wave number-intensity spectra plots in the fetch limited region south of the front (Figure 3a) and beyond the continental shelf (Figure 3b). Both plots show mainly range travelling wave systems coming from the southwest. In the fetch-limited area (Figure 3a), there are three main wave systems: the two long wavelength systems (180 and 150 m) are 30 degrees apart, and the shorter wind-driven wave has a wavelength of 120 m. In the open ocean (Figure 3b), there is one long wavelength system between 185 and 270 m, and a shorter one between 95 and 120 m. These two systems are range travelling waves.

5.2 THE GEOSAT ALTIMETER WAVE ANALYSIS

The Geosat H1/3 and minimum significant swell heights are presented in Figure 4 along with the position of the atmospheric front. The H1/3 variations show a steady increase from the start of the pass to almost the end of the continental shelf from 3 m to 4.5 m. Beyond the continental shelf, the mean H1/3 remains stable at 4.5 m.

The altimeter-derived swell is lower near the Norwegian coast at the start of the pass with a mean height of 2.5 m. North of the atmospheric front the mean swell height is 3.5 m. South of the atmospheric front, where the highest winds were measured, the swell cannot be derived from the altimeter H1/3 and U measurements. The Pierson Moskowitz expression used to derive the swell height from the wind speed measurement assumes that the wind sea is fully developed. In the middle of a storm when the conditions are not fully developed (as is the case here), the altimeter-derived swell height cannot be estimated.

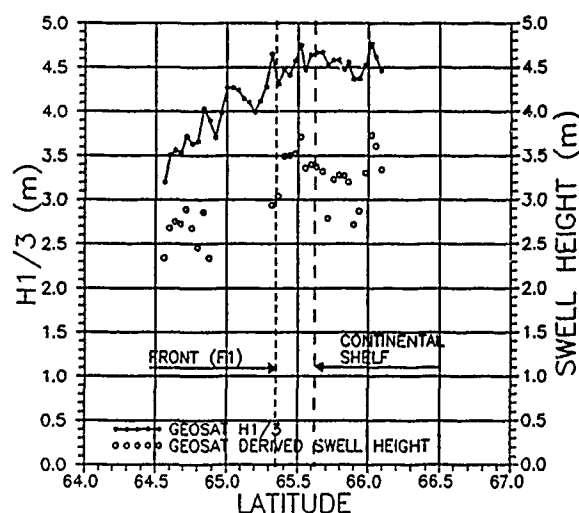


Figure 4: Geosat altimeter H1/3 and swell heights variations on March 20, 1988, over the NORCSEX-88 area.

During the March 20 SAR flight, the dominant swell and wind wave fields observed in the SAR imagery came from the southwest. The four buoys measured a swell coming from the southwest that was still present and dominant during the SAR flight. On March 15, this swell family was observed west of Scotland by the Geosat altimeter, which measured high sea state conditions along one pass with H1/3 of 7.30 m and wind speed of 16.53 m/s at 54.13°N latitude and 17.87°W longitude. Using the JONSWAP parametric formulation for wave generation in the midst of a storm (Hasselmann et al., 1973), the peak period of the waves generated during this storm is estimated at 12.5 s which corresponds to a peak wavelength of 240 m (Mognard et al., 1986). Given the time of observation by Geosat, this swell system should have reached the Haltenbanken area on March 17, and would thus on March 20 only be a residual system arriving from the southwest.

On March 19, at 5 hours UTC, a storm was measured by the Geosat altimeter north of Haltenbanken at latitude 72.04°N, longitude 4.55°E, with maximum H1/3 of 5.38 m and wind speed of 17.64 m/s. Using the JONSWAP parametric wave spectrum expression, the estimate of the wave peak period generated by this storm is 10.30 s and the time of arrival at buoy 4 is March 20 at 8 hours UTC.

5.3 THE BUOY WAVE DATA

An analysis of the buoy data on March 20 (Barstow and Bjerken, 1988) shows that buoy 2 was measuring a local fetch-limited sea with westerly wave direction. Long waves were refracted by the Haltenbanken plateau towards buoy 2, which occupies a sheltered location behind the bank. At 8 hours UTC, the H1/3 measured by buoy 2 was 2.1 m, which is low compared to the altimeter measurement of 3.2 m at the start of the Geosat pass (Figure 4). This 1 m difference between Geosat and buoy 2 might be due to the sheltered location of buoy 2 (at the same time buoy 3 measured 4.7 m and buoy 1 measured 3.6 m at 9 hours UTC, both in good agreement with Geosat). This sheltering effect might explain the almost constant overestimation of Geosat H1/3 in the vicinity of buoy 2 during NORCSEX-88 (Barstow and Bjerken, 1988).

At 9 hours UTC buoy 4 measured an H1/3 of 5.3 m and a bimodal wave system, with a residual southwesterly swell with a 12.5 s period and a northwesterly system with a peak spectral estimate at around 8.5 s. The northwesterly wave system reached the other buoys later during the day and was not imaged by the SAR at the time of the flight. However, the time of arrival of the northwesterly system is in very good agreement with the Geosat estimation of the 10.3 s northerly swell.

6. CONCLUSION

This is the first time that quasi-simultaneous airborne SAR and Geosat altimeter observations were acquired. On March 20 an occluded front was located parallel to the Norwegian coast and was imaged on each of the four SAR flight lines. The observations of this front are characterized on both the airborne SAR and the Geosat altimeter profiles by a sharp variation in the intensity of the backscattered energy. The variation of the SAR normalized backscatter intensity and the altimeter cross section reveal the same features: a fetch-limited region in the vicinity of the Norwegian coast near buoy 2; an atmospheric front at 65.35°N near buoy 4; and a region beyond the continental shelf north of 65.6°N along line 1 where the SAR backscatter was uniform and minor variations in the Geosat cross-section occurred.

The sea state conditions during the simultaneous SAR/Geosat observations were quite complex. The wave conditions measured by the four buoys show the existence of several wave systems:

- 1) a residual southwesterly swell with a 12.5 s peak period that corresponds on the SAR to well-imaged range travelling waves and can be traced by Geosat to a storm that occurred southwest of Scotland;
- 2) a southwest wind-driven wave family, also imaged on the SAR; and

- 3) a new incoming northwesterly swell with a period of 8.5 s that was not imaged on the SAR but that was forecasted using the Geosat altimeter observations.

7. REFERENCES

- Barrick, D.E., Wind dependence of quasi-specular microwave sea scatter, *IEEE Trans. Antennas Propag.*, AP-22, 135-136, 1974.
- Barstow, S.F., and S. Bjerken, Wave measurements from moored directional wave buoys during NORCSEX'88, *ODAP Rep.*, 108, 1988.
- Brown, G.S., et al., The wind speed measurement capability of spaceborne radar altimeters, *IEEE J. Oceanic Eng.*, 6, 1981.
- Dobson, E., et al., Validation of Geosat altimeter-derived wind speeds and significant wave heights using buoy data, *J. Geophys. Res.*, 92, C10, 10719-10731, 1987.
- Hasselmann, K., et al., Measurements of wind-wave growth and swell during the Joint North Sea Wave Project (JONSWAP), *Dtsch. Hydrogr. Z. Suppl.*, A-8, 12, 1973.
- Livingstone, C.E., et al., CCRS C-band airborne radar system: system description and test results, *Proc. 11th Canadian Symp. on Remote Sensing*, Waterloo, Ontario, 503-518, 1987.
- Livingstone, C.E., et al., CCRS C/X airborne Synthetic Aperture Radar: an R&D tool for the ERS-1 timeframe, *IEEE Aerospace and Electronic Syst. Mag.*, 3, 10, 11-20, 1988.
- MacArthur, J.L., et al., The Geosat altimeter, *Tech. Dig. Johns Hopkins APL*, 8, 2, 176-181, 1987.
- Mognard, N.M., and B. Lago, The computation of wind speed and wave height from Geos-3 data, *J. Geophys. Res.*, 84, B8, 3979-3986, 1979.
- Mognard, N.M., Ocean wave parameters extraction using satellite short-pulse radar altimeter, *Proc. of a Workshop on ERS-1 Radar Alt. Data Prod.*, ESA SP-221, 37-41, Frascati, Italy, 1984.
- Mognard, N.M., et al., Southern ocean waves and winds derived from Seasat altimeter measurements, *Wave Dynamics and Radio Probing of the Ocean Surface*, eds O.M. Phillips and K. Hasselmann, 32, 479-489, 1986.
- Monaldo, F.M., and D.R. Lyzenga, On the estimation of wave slope and height variance spectra from SAR imagery, *IEEE Trans. on Geosc. and Remote Sensing*, GE-24, 4, 1986.

GEOSAT DERIVED WINDS, WAVES, AND SWELLS IN THE NORTH ATLANTIC DURING NORCSEX-88

Nelly M. Mognard
Centre National d'Etudes Spatiales
18 Avenue E. Belin
31055 Toulouse-Cedex France

William J. Campbell
Edward G. Josberger
U.S. Geological Survey
University of Puget Sound
Tacoma, WA 98416

ABSTRACT

NORCSEX-88 (Norwegian Continental Shelf Experiment) took place during March 1988 in the Haltenbanken region off the Norwegian coast. The sea state conditions were highly variable, ranging from calm to swell dominated, wind-wave dominated, and mixed swell and wind-wave situations. The Geosat altimeter observed sea surface wind speeds and significant wave heights over the Northeast Atlantic during the NORCSEX-88 experiment. These observations were used to derive fields of minimum significant swell heights and maximum significant wind waves, and also parametric Jonswap wave spectra.

The averaged monthly mean fields of wind speed, wave height, and swell heights for March 1988 show that the sea state conditions in the Haltenbanken region of the NORCSEX-88 experiment were not extreme when compared to the conditions in the North Atlantic during this period. The three-day averaged mean sea state fields, from the Geosat altimeter, show the evolution and propagation of wind and wave patterns across the North Atlantic. A swell family, created off Newfoundland, propagated all the way across the North Atlantic to the Haltenbanken region. Individual Geosat passes across two particularly intense storms during the NORCSEX-88 experiment are analyzed and compared to weather maps, wave hindcasts, and wavescan buoy measurements. These comparisons show that the Geosat altimeter can accurately locate atmospheric fronts, determine swell propagation, and give an estimate of parametric wave spectra for wind-wave dominated situations.

Key words: Geosat, NORCSEX, H1/3, Wind Speed, Swell

1. INTRODUCTION

Over the ocean, the Geosat radar altimeter measures the significant wave height (H1/3) and the surface wind speed (U) over a distance of approximately 7 km along the satellite track. The H1/3 measurement has an accuracy of 50 cm or 10 percent of H1/3, whichever is greater (Dobson et al., 1987). The wind speed is determined from the amount of power reflected by the ocean surface using the algorithm developed for the Geos-3 altimeter (Brown, 1981). The wind speed measurements have an accuracy of 1.8 m/s for winds less than 18 m/s (Dobson et al., 1987). For winds higher than 18 m/s, the power

reflected toward the altimeter antenna is greatly reduced and the precision of the wind speed measurement is difficult to estimate (Mognard and Lago, 1979).

An estimate of minimum significant swell height can be deduced from the altimeter sea state measurements (H1/3 and U), except within ocean areas covered by intense storms (Mognard et al., 1986). Along an altimeter track the maximum height of the locally generated wind waves, corresponding to fully developed conditions, can be determined from the altimeter wind speed measurements using the Pierson and Moskowitz formulation (Pierson and Moskowitz, 1962). If the altimeter H1/3 is higher than the estimated fully developed wind wave height, then a minimum significant swell height can be computed (Mognard, 1984). This technique, developed using Seasat altimeter data, has tracked swell propagating for several days across the South Pacific (Mognard, 1984) and the North Atlantic (Mognard, 1983).

Using the global data set acquired by Seasat altimeter during its 3-month lifetime (July to October 1978), monthly fields of wind speed, H1/3, and swell revealed the presence of mesoscale features that are not present in the long-term monthly climatological fields (Mognard et al., 1983). This same technique has been used to compute the altimeter-derived monthly fields in the North Atlantic during NORCSEX-88 in March 1988 (Figure 1).

Sea state fields derived from the Geosat altimeter in the North Atlantic for two consecutive 3-day periods (March 6-8 and 9-11) are shown in Figures 2, 3, and 4. During these 6 days a swell family generated during the March 6-8 period in the Northwest Atlantic off Newfoundland propagated across the Atlantic and reached the Haltenbanken region on March 11 during a local storm. The Northwest Atlantic storm responsible for generating the swell was sampled by the altimeter on March 6 and 7. For those 2 days, a JONSWAP parametric spectrum is used to determine the wind-wave peak frequency. The arrival time of the resulting swell in the NORCSEX area is then estimated.

A Geosat pass in the midst of the local storm that occurred on March 11 in the Haltenbanken region revealed the presence of a 3-5 m background swell superimposed on the locally generated wind waves. Comparisons between the Geosat estimates of swell height and period and the Norwegian WINCH hindcasts

swell give a very good agreement for the height, period, and direction of the swell. Geosat H1/3 and wind speed measurements in the center of the March 11 storm are in good agreement with simultaneous buoy observations.

II. GEOSAT MEAN FIELDS IN THE NORTH ATLANTIC FOR MARCH 1988

Monthly fields of wind speed, H1/3, and swell heights, characterize the regional sea state features of the North Atlantic during the NORCSEX-88 experiment. Various sea state features which appear in the long-term climatological fields, such as the region of high sea state in the Northwest Atlantic, are observed in the Geosat altimeter averaged monthly fields.

II.1 THE WIND SPEED FIELD

The Geosat averaged wind speed field during March 1988 is characterized by maximum winds of 9.25 m/s and minimum winds of 7.25 m/s (Figure 1a). The region of minimum wind speed is located in the Northeast Atlantic along the Norwegian coast. The Haltenbanken area had average winds between 7.25 m/s and 7.50 m/s, which is low compared to the rest of the North Atlantic. There are two regions of maximum wind speed: a large region of averaged winds higher than 9 m/s in the middle of the North Atlantic, south of Iceland, and another region in the western part of the North Atlantic between Greenland and Newfoundland (partly shown in Figure 1a) where intense storms usually occur during boreal winters.

II.2 THE H1/3 FIELD

The mean H1/3 field obtained with the Geosat altimeter in the North Atlantic in March 1988 is shown in Figure 1b. The western part of the North Atlantic is covered with H1/3 ranging between 3 and 3.5 m, while north and east of a line from Iceland to Scotland, there is a gradual decrease of H1/3 from 3 to 2 m. The maximum averaged H1/3 is 3.5 m and the minimum is 2 m. The Haltenbanken region has average H1/3 of 2.25 m.

II.3 THE SWELL FIELD

During March 1988, the mean significant swell height, derived from the Geosat altimeter, vary between 2.25 and 0.75 m (Figure 1c). The highest swells, greater than 2 m, are located South of 50°N and West of 10°W. North of 50°N, the mean swell height uniformly decreases to 0.75 m, north of 70°N. The Haltenbanken region has a mean swell of 1.25 m slightly but significantly higher than the rest of the averaged swell in March along the Norwegian coasts.

In summary, during March 1988, the mean monthly sea state fields derived from the Geosat altimeter show that the highest sea state were located in the western part of the North Atlantic, the Northeast Atlantic being characterized with lower averaged sea state values.

NORTH ATLANTIC MARCH 1988 -WIND SPEED(m/s)-



NORTH ATLANTIC MARCH 1988 -H1/3(m)-



NORTH ATLANTIC MARCH 1988 -SWELL(m)-

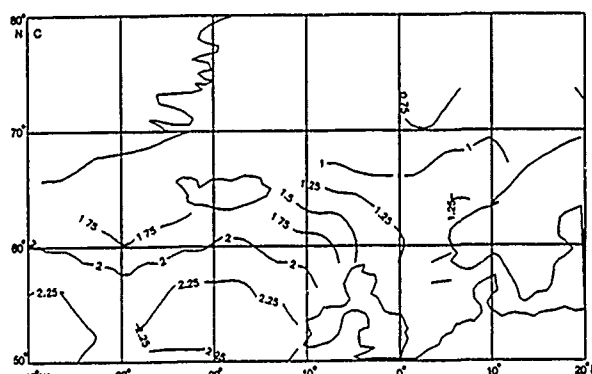


Figure 1: Geosat mean monthly fields of wind speed (a), H1/3 (b), and swell height (c) in the North Atlantic during March 1988.

III. GEOSAT 3-DAY AVERAGED FIELDS IN THE NORTH ATLANTIC

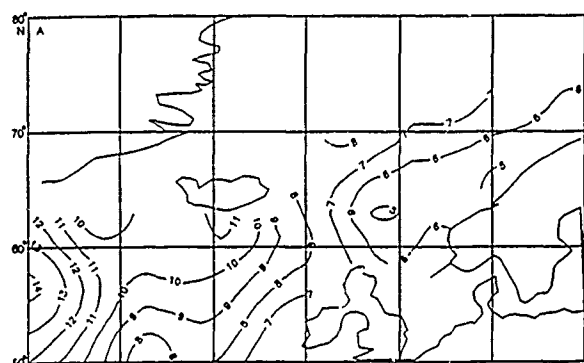
The Geosat 3-day fields present a quasi-synoptic view of the ocean sea state of the North Atlantic. Mognard et al. (1986) found that given the sampling spatial coverage of the Seasat radar altimeter, a 3-day averaging of all observations in a given ocean area provided sufficiently spaced data to generate a quasi-synoptic view of rapidly changing conditions in the area. These fields delineate areas where intense storms have occurred and show the propagation of ocean swell. Two consecutive 3-day periods during the beginning of March 1988 are presented: March 6-8, and March 9-11. The first

3-day period was a very calm period in the Haltenbanken region with an intense storm in the Northwest Atlantic. This storm, south of Greenland and east of Newfoundland, generated a swell family that reached the NORCSEX area after 3 to 4 days. The second 3-day period from March 9-11, had a storm in the Northeast Atlantic where the swell generated during the first 3-day period was superimposed on the wind wave sea created by the ongoing storm.

III.1 THE 3-DAY WIND SPEED FIELDS

During March 6-8, intense winds with an average maximum of 14 m/s occurred South of Greenland (Figure 2a). Averaged winds less than 10 m/s occurred in the eastern part of the North Atlantic ocean. The lowest mean winds were observed along the Norwegian coast and in the NORCSEX area where they were less than 6 m/s. During the next 3-day period, from March 9-11 (Figure 2b), the highest mean winds of 11 m/s were located in the Northeast Atlantic, along the Norwegian coast and in the NORCSEX area. However, the averaged winds were lower than in the intense storm in the Northwest Atlantic of the preceding 3-day period.

6 to 8 March 1988 -WIND SPEED(m/s)-



9 to 11 March 1988 -Wind Speed(m/s)-



Figure 2: Geosat 3-day wind speed fields in the North Atlantic for (a) March 6-8 and (b) March 9-11, 1988.

III.2 THE 3-DAY H1/3 FIELDS

During the March 6-8 period, the average H1/3 field measured by the Geosat altimeter (Figure 3a) shows that the highest magnitude, 6 m, occurred south of Greenland where the highest winds were observed during the same period (Figure 2a). A large region

with H1/3 ranging from 3-6 m occurred in the Northwest Atlantic, with a gradual decrease in H1/3 in the Northeast Atlantic where the Norwegian coasts and the NORCSEX area had H1/3 less than 1.5 m. This situation changes completely in the next 3-day period (Figure 3b) where two regions with waves greater than 4 m are found in the mid-Atlantic extending from Iceland to Scotland, and in the Haltenbanken region. These two regions of high H1/3 are composed of swell and wind waves, as can be seen on the 3-day swell fields.

6 to 8 March 1988 -H1/3(m)-



9 to 11 March 1988 -H1/3(m)-

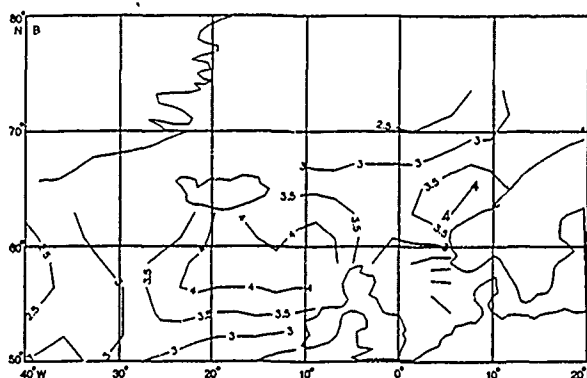


Figure 3: Geosat 3-day H1/3 fields in the North Atlantic for (a) March 6-8 and (b) March 9-11, 1988.

III.3 THE 3-DAY SWELL FIELDS

The averaged 3-day swell fields (Figures 4a-b) show the swell fields which occurred in the North Atlantic. The first 3-day field, the period March 6-8 (Figure 4a), shows an area of high swell, with a maximum of 3.5 m, in the Northwest Atlantic, south of Iceland. This shows the start of the swell propagation from the region of high winds and high H1/3 observed during the same period south of Greenland (Figures 2a and 3a). During this first 3-day period, northeast of a line between Iceland and Scotland, the swell was less than 2 m, and less than 1 m in the NORCSEX region. Three days later (Figure 4b) the core of swell, higher than 3 m, moved toward the east and extended between Iceland and Scotland. The 2 m swell moved towards the northeast and reached the Haltenbanken area, where mixed south westerly swell was superimposed on a wind sea generated by a storm on March 10 and 11.

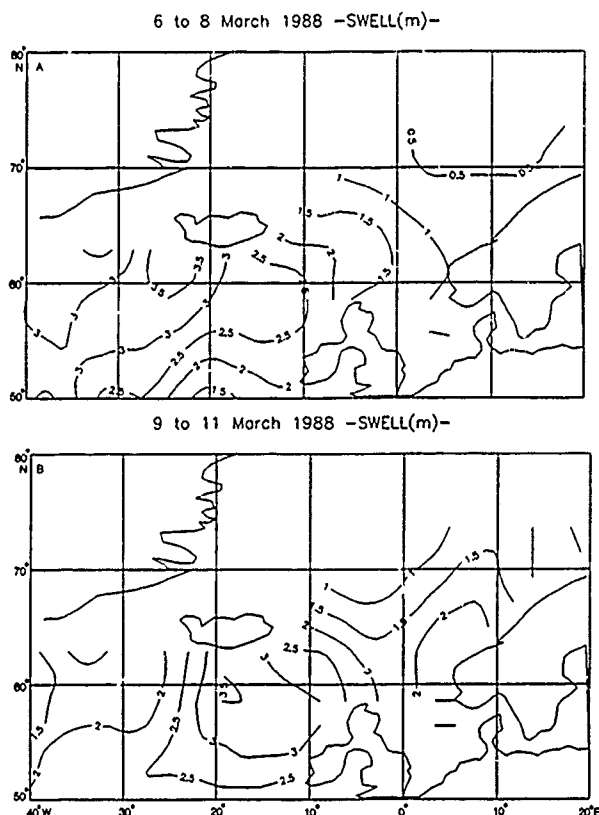


Figure 4: Geosat 3-day derived swell height fields in the North Atlantic for (a) March 6-8 and (b) March 9-11, 1988.

IV. THE SWELL GENERATION AREA IN THE NORTHWEST ATLANTIC

IV.1 THE JONSWAP SPECTRA IN THE MIDST OF THE STORM

Between March 6 and 8 an intense storm occurred in the Northwest Atlantic. Parts of this storm were observed by the Geosat altimeter on two consecutive days. On March 6 at 12 hours UTC, Geosat measured high winds along one pass with maximum wind speed of 20 m/s and H1/3 of 7 m at 56°N and 50°W. On March 7, at 3 hours UTC, the Geosat altimeter measured H1/3, varying between 5 and 9 m, with winds reaching 13 m/s at 60°N and 40°W. This storm then moved northeast.

In the absence of swell, H1/3 measured by the altimeter in the midst of a storm is representative of either a fully developed condition or a duration-limited situation. In this case, nearly all spectra can be determined by a JONSWAP parametric form of the spectrum (Hasselmann et al., 1973). This parametric representation was used with the Seasat altimeter to estimate wave spectra in the Southern Ocean during winter storm conditions (Mognard et al., 1986).

The peak frequency of the JONSWAP spectrum (f_m) can be determined by means of the altimeter measured wind speed and H1/3 using the JONSWAP dimensionless energy frequency relationship (Hasselmann et al., 1976). For March 6, the observed wind of 20 m/s and H1/3 of 7 m gives a peak frequency $f_m = 0.085$ Hz, or a

period of 11.7 s. For March 7, the measured maximum wind speed of 18 m/s corresponds to an H1/3 of 7.5 m and gives a peak frequency of 0.079 Hz, or a period of 12.6 s.

IV.2 THE SWELL PROPAGATION TO THE HALTENBANKEN REGION

The two Geosat passes of March 6 and 7 in the Northwest Atlantic may not have observed the highest wind speeds and H1/3 that occurred in this particular storm. Indeed, given the sparse temporal and spatial coverage, they probably did not. The sea state conditions measured on March 7 are representative of more fully developed conditions than the sea state parameters measured on March 6. Considering the sea state conditions measured on March 7, the propagation velocity of waves with a 12.6 s period is 10 m/s. These waves should arrive in the NORCSEX area approximately 3.13 days after the time when they were generated, on March 10 at 11 hours UTC. If we assume that the highest sea state in this storm could occur only during a period within 6 hours before or after the time when Geosat measured the highest sea state, it gives a propagation time for the 12.6 s period waves between 2.88 and 3.38 days. It is impossible to evaluate what the highest sea state in this storm was without any other data acquired at a different time or place than the Geosat passes. Therefore, the 12.6 s period is only indicative of the period of the swell propagating from the storm. If the Geosat missed the highest sea state conditions, then the period of the resulting swell would have been greater.

V. THE MARCH 11 STORM IN THE NORCSEX AREA

The Geosat altimeter at 1 hour UTC on March 11 acquired data during an intense storm in the NORCSEX area. Figure 5 shows the variations of H1/3 and wind speed along the pass and the quasi-simultaneous measurements at 0 and 3 hours UTC of a buoy located along the satellite track. The sharp change in wind speed at 65°N latitude from 5 to 15 m/s corresponds to the crossing of an occluded front. East of this front where the altimeter winds are low, between 5 and 6 m/s, the H1/3 measured by the altimeter is about 3 m, indicating the presence of a swell. Thus, the wave regime along the pass is a combination of wind wave and swell. Using the Pierson-Moskowitz formulation for fully developed waves, minimum significant swell heights are estimated and compared to the WINCH hindcasts along the satellite track (Figure 6). In the middle of the storm, where the H1/3 measured by the altimeter is lower than the fully developed wind wave, no swell heights can be estimated. The WINCH swell periods along the track vary between 13.1 s and 12.8 s, with a mean propagation direction toward the east. The swell direction is in perfect agreement with the Geosat location of the generating storm in the Northwest Atlantic. The Geosat estimate of the swell peak period at 12.6 s indicates that the Geosat tracks came close to the center of the storm and the altimeter observed nearly the maximum sea state that occurred.

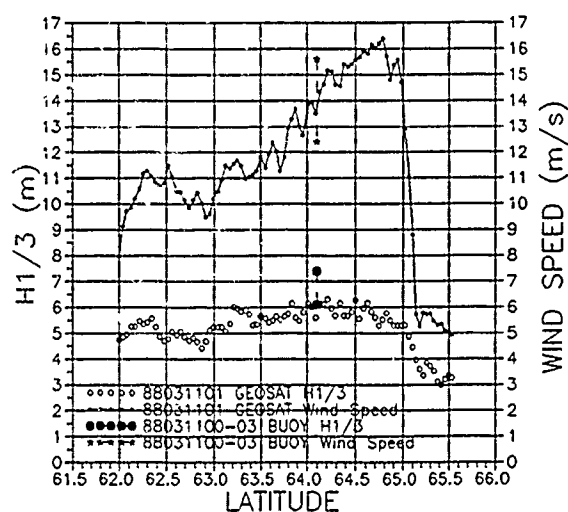


Figure 5: Geosat wind speed and H1/3 variations at 1 hour UTC during the March 11 storm in the NORCSEX region, with buoy measurements along the track at 0 and 3 hours UTC.

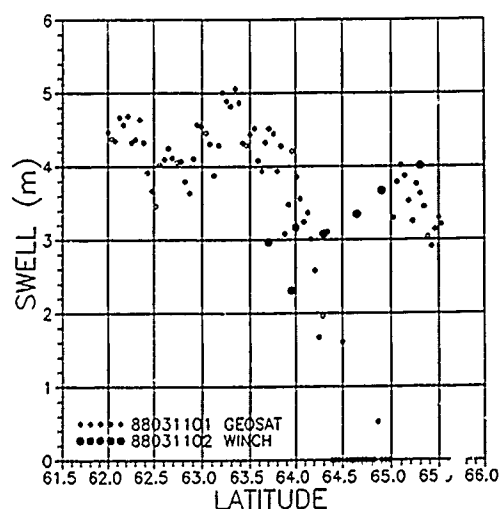


Figure 6: Geosat deduced swell heights along the pass on March 11 in the NORCSEX area and WINCH swell heights hindcasts along the track.

VI. REFERENCES

- Brown, G.S., et al., The wind speed measurement capability of spaceborne radar altimeters, *IEEE J. Oceanic Eng.*, 6, 1981.
- Dobson, E., et al., Validation of Geosat altimeter-derived wind speeds and significant wave heights using buoy data, *J. Geophys. Res.*, 92, C10, 10719-10731, 1987.
- Hasselmann, K., et al., Measurements of mid-wave growth and swell during the Joint North Sea Wave Project (JONSWAP), *Dtsch. Hydrogr. Z., Suppl.* A8 (12), 1973.
- Hasselmann, K., et al., A parametric wave prediction model, *J. Phys. Oceanogr.*, 6, 200-228, 1976.
- Mognard, N.M., and B. Lago, 1979, The computation of wind speed and wave height from Geos-3 data, *J. Geophys. Res.*, 84, B8, 3979-3986, 1979.
- Mognard, N.M., Swell propagation in the North Atlantic ocean using Seasat altimeter, *Satellite Microwave Remote Sensing*, ed. T.D. Allen, 425-438, 1983.
- Mognard, N.M., et al., Southern ocean monthly waves and surface winds for winter 1978 by Seasat radar altimeter, *J. Geophys. Res.*, 88, C3, 1736-1744, 1983.
- Mognard, N.M., Swell in the Pacific ocean observed by Seasat radar altimeter, *Marine Geodesy*, 8, 183-209, 1984.
- Mognard, N.M., et al., Southern ocean waves and winds derived from Seasat altimeter measurements, *Wave Dynamics and Radio Probing of the Ocean Surface*, eds. O.M. Phillips and K. Hasselmann, 32, 479-489, 1986.
- Pierson, W.J., and L. Moskowitz, A proposed spectral form for fully developed wind seas based on the similarity theory of S.A. Kitaigorodskii, *J. Geophys. Res.*, 69, (24), 5181-5190, 1963.

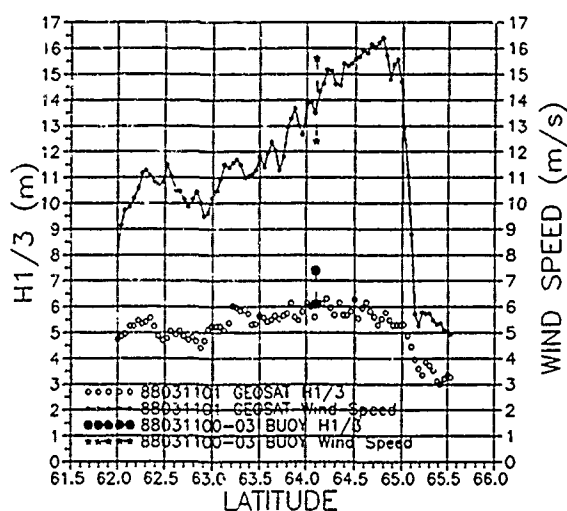


Figure 5: Geosat wind speed and $H1/3$ variations at 1 hour UTC during the March 11 storm in the NORCSEX region, with buoy measurements along the track at 0 and 3 hours UTC.

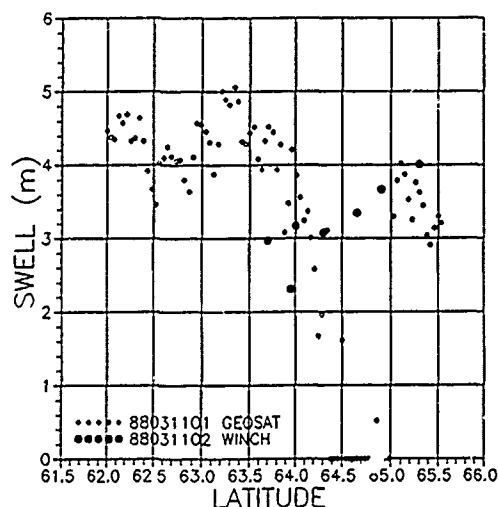


Figure 6: Geosat deduced swell heights along the pass on March 11 in the NORCSEX area and WINCH swell heights hindcasts along the track.

VI. REFERENCES

- Brown, G.S., et al., The wind speed measurement capability of spaceborne radar altimeters, *IEEE J. Oceanic Eng.*, 6, 1981.
- Dobson, E., et al., Validation of Geosat altimeter-derived wind speeds and significant wave heights using buoy data, *J. Geophys. Res.*, 92, C10, 10719-10731, 1987.
- Hasselmann, K., et al., Measurements of mid-wave growth and swell during the Joint North Sea Wave Project (JONSWAP), *Dtsch. Hydrogr. Z., Suppl. A8* (12), 1973.
- Hasselmann, K., et al., A parametric wave prediction model, *J. Phys. Oceanogr.*, 6, 200-228, 1976.

Mognard, N.M., and B. Lago, 1979, The computation of wind speed and wave height from Geos-3 data, *J. Geophys. Res.*, 84, B8, 3979-3986, 1979.

Mognard, N.M., Swell propagation in the North Atlantic ocean using Seasat altimeter, *Satellite Microwave Remote Sensing*, ed. T.D. Allen, 425-438, 1983.

Mognard, N.M., et al., Southern ocean monthly waves and surface winds for winter 1978 by Seasat radar altimeter, *J. Geophys. Res.*, 83, C3, 1736-1744, 1983.

Mognard, N.M., Swell in the Pacific ocean observed by Seasat radar altimeter, *Marine Geodesy*, 8, 183-209, 1984.

Mognard, N.M., et al., Southern ocean waves and winds derived from Seasat altimeter measurements, *Wave Dynamics and Radio Probing of the Ocean Surface*, eds. O.M. Phillips and K. Hasselmann, 32, 479-489, 1986.

Pierson, W.J., and L. Moskowitz, A proposed spectral form for fully developed wind seas based on the similarity theory of S.A. Kitaigorodskii, *J. Geophys. Res.*, 69, (24), 5181-5190, 1963.

A Study of Textural and Tonal Information for Classifying Sea Ice from SAR Imagery

T.K. Hirose

L. McNutt

J.S. Paterson

Noetix Research Inc.
902-280 Albert St.
Ottawa, Ontario

Canada Centre for R.S.
200-110 O'Connor St.
Ottawa, Ontario

Intera Technologies Ltd.
200-110 O'Connor St.
Ottawa, Ontario

Abstract

Many textural and tonal measures have been proposed and used to classify sea ice from imagery collected by a synthetic aperture radar (SAR). These measures, however, are computationally expensive to calculate. Moreover, the behaviour of some measures in the sea ice application is largely unknown.

In this study, the characteristics and relationship between five textural measures derived from the spatial grey levels or co-occurrence matrix, non-coherent averaging, standard deviation, and a texture measure that accounts for speckle noise are examined for a X-band and L-band SAR dataset. Initial results suggest that a high correlation exists between measures and therefore, significant savings in computation can be achieved by eliminating redundant channels.

1.0 Introduction

With the launch of E-ERS-1 and RADARSAT satellites, a large volume of data will be available. The manual extraction of the location and identification of ice types, however, is slow and may not meet the throughput requirements for operational applications. To improve turnaround, automatic techniques to classify ice types could be implemented.

It is well known that both tone and texture can be used to discriminate between ice types. However, the characteristics and utility of different textural measures for the sea ice application is not well understood.

In this paper, the characteristics of some tone and textural features are investigated. In section two, the datasets used in this study are described while section three outlines the procedure used to process the data and extract statistical information on the sea ice. The analysis of the statistical data is discussed in section four and conclusions follow in section five.

2.0 Dataset Description

Two digital SAR datasets are evaluated in this study. The Mould Bay scene was acquired with the Intera STAR-1 system and the second, a view of the marginal ice zone in the Beaufort

Sea, was imaged with the SEASAT satellite. The system configuration for each sensor is summarized in Table 1.

In conjunction with the SAR pass over Mould Bay on March 3, 1984, a surface truth program was started on March 15 (Bjerkelund et. al., 1984). Although a two week period separated the overflight and surface truth collection, the datasets are comparable because the ice remained stable during this period (Cameron M.A., 1986). In the scene, a cluster of old ice floes are in the middle of the bay and surrounded by thinner first-year (FY) ice. Landfast ice is present along the shore and in a sheltered bay.

Prior to processing the digital imagery for Mould Bay, the resolution and size of the imagery was reduced by a factor of two. This was achieved by convolving the image with a 2 by 2 averaging filter then decimating by two in both the x and y directions to produce an image one quarter the area of the original, or 12 meter square pixels.

The second dataset, collected by the SEASAT satellite on October 5, 1978, covers a 100 km square area of sea ice in the marginal ice zone. Over 60 percent of the scene is covered with old ice floes. Young FY ice covers the remaining portion of the scene except where new ice has recently formed in areas of floe movement.

No preprocessing of the SEASAT data was performed before input to the processing stage described in the next section.

3.0 Data Processing

The data processing was performed on an image analysis system (IAS) manufactured by DIPIX Technologies Ltd. using the standard software package as well as custom software developed at the Canada Centre for Remote Sensing.

In the first stage, tone and textural features were generated for each location on the image. Then, representative samples of the different ice types were manually delineated on the IAS. Statistics of the samples for the features were calculated for each ice type and graphed as illustrated in Figures 1 and 2. To supplement the graphs, the correlation between features for each ice type was determined and tabulated (Tables 2 and 3).

The nine features evaluated in this study include :

correlated (Tables 3a-c) but confusion between new and young ice types exist.

Contrast, dissimilarity, and the standard deviation values (Figure 3b), show new and young ice with overlapping signatures. Commensurate with the other features, old ice is separable from new and young ice.

The field texture shows a very small signature variation for new ice. However, the new ice signature is within the range of the young ice values indicating confusion between the two ice types. Old ice has significantly lower values and is separable from the others.

In summary, two classes of ice are clearly distinguishable based on the average values for the individual features: new-young ice, and old ice. It is unknown with this analysis, however, if the features used in combination would discriminate between the new and young ice types.

Mould Bay vs Marginal Ice Zone

When highly correlated features in one scene are compared with correlated features in the other, a consistent behaviour is observed. Correlated features are the same in both scenes. Moreover, four distinct groups containing correlated features can be recognised.

In the first group, the tone and mean filtered image are correlated. The reason for this is obvious. Entropy, homogeneity, and uniformity form the second group with correlation values greater than 0.8. Similar correlation values (> 0.8) are observed within the third group which consists of contrast, dissimilarity, and the standard deviation measures. Field texture is in a group of itself with little correlation with the others.

The dissimilarity and standard deviation measures are also correlated with the entropy, homogeneity, and uniformity group, especially for first year ice types.

5.0 Conclusions

In this study, the characteristics of nine spectral and textural features are investigated. It was found that they could be divided into four groups, each group containing correlated features. One group contains the tone and mean filtered images, a second with entropy, homogeneity, and uniformity measures. A third group consists of contrast, dissimilarity, and standard deviation. The field texture measure is in the fourth group.

These initial results suggest that some of the measures are redundant and could be eliminated from a classification scheme without affecting the results. However, the selection of best features to be used alone or in combination is a subject for future research. Further study on the affect of different distance separations and optimal window size used to generate the co-occurrence matrix is also necessary.

Acknowledgements

We would like to acknowledge the Canada Centre for Remote Sensing, Dept. of Energy, Mines, and Resources for funding this project. We are also grateful to the Canadian Atmospheric Environment Service (Ice Branch) for making the X-band data available and to the Jet Propulsion Laboratory in Pasadena, California for the SEASAT data.

References

- Bjerkelund, C., D. Lapp, and S. Prashker "An investigation and Report on the Definition of the Physical Properties of Ice Types at Mould Bay and Isachsen, N.W.T.". 1984. Volumes 1 and 2. Prepared for AES. Vol. 1 128 pp. and Vol. 2 106 pp.
- Cameron, M.A. "Analysis of Winter Ice Using Digital STAR-1 SAR Imagery of Mould Bay, N.W.T.". 1986. Master's Thesis prepared for the Dept. of Geography, Carleton University. 152 pp.
- Burns B.A., and D.R. Lyzenga. "Textural Analysis as a SAR Classification Tool". 1984. Elec. vol 4, pp. 309-322.
- Haralick, R.M.. "Statistical and Structural Approaches to Texture". 1979. Proc. IEEE, vol. 67, pp. 786-804.
- Wezka, J.S., C.R. Dyer and A. Rosenfeld. "A Comparative Study of Texture Measures for Terrain Classification". 1976. IEEE Trans. on Systems, Man, and Cybernetics. Vol. SMC-6, No. 4. pp. 269-285.

Table 1.0. Sensor Characteristics.

| SENSOR | WAVELENGTH POLARIZATION | INCIDENCE NEAR | ANGLE FAR | PIXEL (m) | SWATH WIDTH (km) |
|--------|----------------------------|-------------------|--------------|--------------|------------------------|
| SEASAT | L (23.5 cm) HH | 18 | 25 | 12.5 | 100 |
| STAR-1 | X (3.2 cm) HH | 68.5 | 78.9 | 6 | 23 |

Table 2.0. Mould Bay correlation coefficients between features.

a) Correlation coefficient matrix for Old Ice in Mould Bay.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| DISS | 0.944 | | | | | | | |
| ENTR | 0.302 | 0.496 | | | | | | |
| HOMO | -0.378 | -0.589 | -0.931 | | | | | |
| MEAN | 0.015 | -0.003 | -0.368 | 0.341 | | | | |
| ORIG | 0.071 | 0.054 | -0.272 | 0.238 | 0.829 | | | |
| STDE | 0.926 | 0.957 | 0.442 | -0.519 | 0.015 | 0.069 | | |
| TEXT | 0.714 | 0.665 | 0.289 | -0.336 | -0.276 | -0.132 | 0.749 | |
| UNIF | -0.184 | -0.369 | -0.945 | 0.879 | 0.426 | 0.331 | -0.296 | -0.181 |

b) Correlation coefficient matrix for Landfast ice in Mould Bay.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| DISS | 0.946 | | | | | | | |
| ENTR | 0.718 | 0.826 | | | | | | |
| HOMO | -0.814 | -0.933 | -0.902 | | | | | |
| MEAN | 0.433 | 0.458 | 0.391 | -0.433 | | | | |
| ORIG | 0.341 | 0.381 | 0.318 | -0.358 | 0.829 | | | |
| STDE | 0.908 | 0.933 | 0.770 | -0.848 | 0.479 | 0.381 | | |
| TEXT | 0.328 | 0.298 | 0.192 | -0.207 | -0.028 | -0.017 | 0.352 | |
| UNIF | -0.561 | -0.785 | -0.957 | 0.871 | -0.366 | -0.300 | -0.716 | -0.162 |

c) Correlation coefficient matrix for FY ice in Mould Bay.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| DISS | 0.934 | | | | | | | |
| ENTR | 0.609 | 0.758 | | | | | | |
| HOMO | -0.701 | -0.882 | -0.860 | | | | | |
| MEAN | 0.364 | 0.419 | 0.356 | -0.412 | | | | |
| ORIG | 0.341 | 0.392 | 0.329 | -0.384 | 0.929 | | | |
| STDE | 0.903 | 0.951 | 0.728 | -0.813 | 0.444 | 0.406 | | |
| TEXT | 0.239 | 0.182 | 0.078 | -0.064 | -0.623 | -0.553 | 0.222 | |
| UNIF | -0.477 | -0.647 | -0.946 | 0.823 | -0.312 | -0.287 | -0.606 | -0.034 |

Table 3.0. Marginal ice zone correlation coefficients between features.

a) Correlation coefficient matrix for New Ice in the marginal ice zone.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| DISS | 0.906 | | | | | | | |
| ENTR | 0.557 | 0.719 | | | | | | |
| HOMO | -0.692 | -0.908 | -0.308 | | | | | |
| MEAN | 0.482 | 0.482 | 0.386 | -0.390 | | | | |
| ORIG | 0.348 | 0.347 | 0.276 | -0.282 | 0.721 | | | |
| STDE | 0.856 | 0.901 | 0.623 | -0.766 | 0.538 | 0.366 | | |
| TEXT | 0.278 | 0.240 | 0.103 | -0.166 | -0.571 | -0.434 | 0.270 | |
| UNIF | -0.421 | -0.613 | -0.931 | 0.773 | -0.306 | -0.218 | -0.498 | -0.070 |

b) Correlation coefficient matrix for Old Ice in the marginal ice zone.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|-------|
| DISS | 0.960 | | | | | | | |
| ENTR | 0.562 | 0.678 | | | | | | |
| HOMO | -0.641 | -0.794 | -0.830 | | | | | |
| MEAN | 0.432 | 0.462 | 0.376 | -0.395 | | | | |
| ORIG | 0.347 | 0.375 | 0.300 | -0.326 | 0.794 | | | |
| STDE | 0.913 | 0.934 | 0.630 | -0.700 | 0.475 | 0.360 | | |
| TEXT | 0.088 | 0.033 | -0.084 | 0.073 | -0.822 | -0.662 | 0.056 | |
| UNIF | -0.430 | -0.556 | -0.928 | 0.806 | -0.315 | -0.249 | -0.497 | 0.093 |

c) Correlation coefficient matrix for Young Ice in the marginal ice zone.

| | CONT | DISS | ENTR | HOMO | MEAN | ORIG | STDE | TEXT |
|------|--------|--------|--------|--------|--------|--------|--------|--------|
| DISS | 0.907 | | | | | | | |
| ENTR | 0.548 | 0.755 | | | | | | |
| HOMO | -0.578 | -0.829 | -0.884 | | | | | |
| MEAN | 0.584 | 0.700 | 0.526 | -0.612 | | | | |
| ORIG | 0.491 | 0.585 | 0.442 | -0.513 | 0.832 | | | |
| STDE | 0.885 | 0.953 | 0.700 | -0.753 | 0.713 | 0.564 | | |
| TEXT | 0.541 | 0.371 | 0.181 | -0.128 | -0.261 | -0.255 | 0.417 | |
| UNIF | -0.392 | -0.606 | -0.921 | 0.845 | -0.419 | -0.346 | -0.541 | -0.104 |

Figure 1a. Mean and standard deviation for ice in Mould Bay.

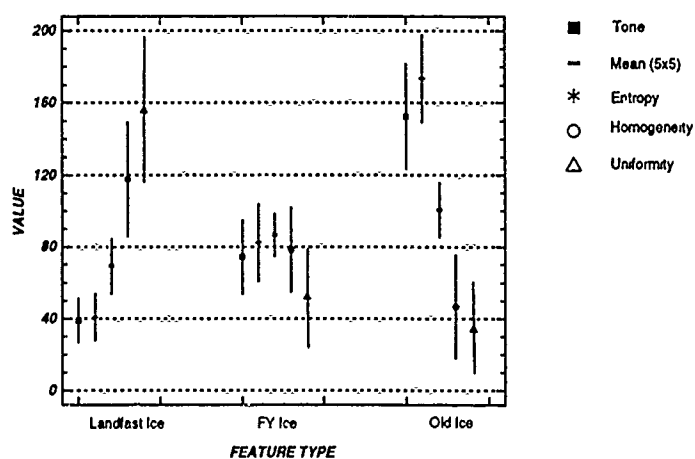


Figure 2a. Mean and standard deviation for ice in the MIZ.

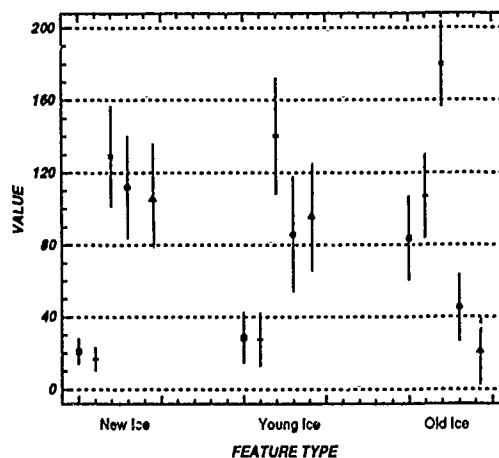


Figure 1b. Mean and standard deviation for ice in Mould Bay.

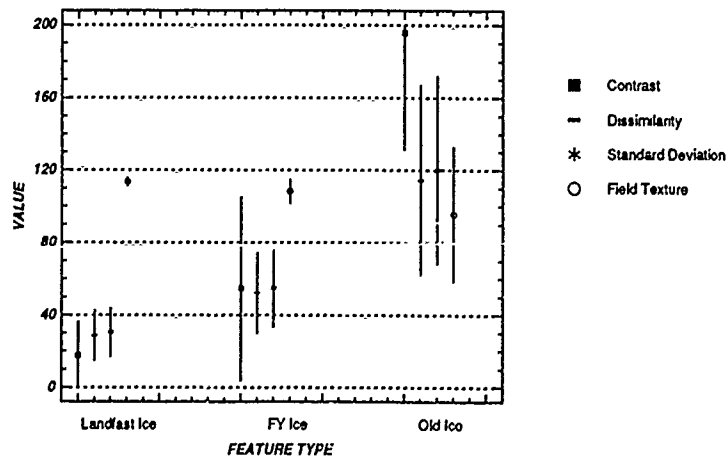
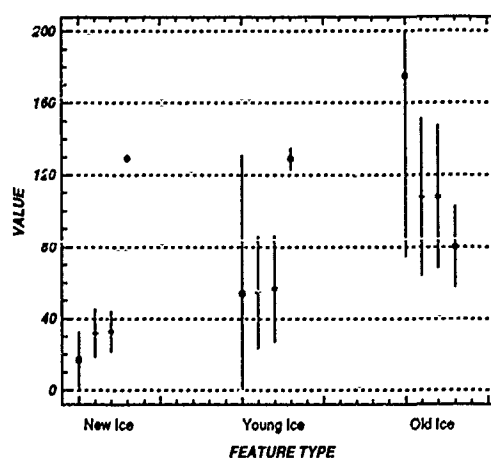


Figure 2b. Mean and standard deviation for ice in the MIZ.



ICE CLASSIFICATION ALGORITHM DEVELOPMENT AND VERIFICATION FOR THE ALASKA SAR FACILITY USING AIRCRAFT IMAGERY

Benjamin Holt, Ronald Kwok and Eric Rignot
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive
Pasadena CA 91109 USA

ABSTRACT

The Alaska SAR Facility (ASF) at the University of Alaska Fairbanks is a NASA program designed to receive, process, and archive SAR data from ERS-1 as well as support science investigations that will use this regional data. As part of ASF, specialized subsystems and algorithms are under development to produce certain geophysical products from the SAR data, in particular ice motion, ice classification, and ice concentration. This paper focuses on the algorithm under development for ice classification and the verification of the algorithm using recently acquired C-band aircraft SAR imagery over the Alaskan Arctic.

Keywords: Ice classification, aircraft SAR, algorithm development.

1. INTRODUCTION

The Geophysical Processing System (GPS) of the Alaska SAR Facility (ASF) is being designed to automatically process ice and ocean geophysical products from SAR data received from the European ERS-1, Japanese ERS-1, and Canadian Radarsat missions. The ASF, located at the University of Alaska Fairbanks, will receive, process, and archive SAR products obtained within its station mask from the three satellites. The GPS will then produce ice motion, ice type, and ice concentration maps as well as limited quantities of ocean directional wave spectra in geophysical units. This paper focuses on the ice classification algorithm under development and some preliminary results using C-band aircraft SAR imagery which is quite similar to the SAR data expected from the European ERS-1 sensor.

The ERS-1 SAR is single frequency (C-band, 5.3 GHz), single polarization (VV) and has a fixed look angle of 23 degrees. The satellite will be in a sun-synchronous, 98 degree orbit and is expected to acquire extensive imagery of sea ice over the Arctic Ocean and adjacent seas. The GPS is being designed to utilize the low resolution (100 m pixel spacing) data product processed at ASF. The low resolution product is an 8 by 8 pixel averaged sea ice made from the 4-look ground corrected full resolution (25 m) image and is 1024 by 1024 pixels in size.

The aircraft SAR imagery used in the algorithm development was obtained during the March 1988 NASA DC-8 SSM/I validation program which obtained SAR imagery over many regions of the Alaskan Beaufort, Chukchi, and Bering Seas. The aircraft SAR, designed and built by JPL, operates at three frequencies, C-, L-, and P-bands with full quad-polarization. The imagery has a resolution in range and azimuth of 10 m and a swath width of 7-10 km over an incidence angle range from 20 to 60 degrees. In addition to the SAR, other sensors which obtained data during the program included the AMMR profiling radiometer and the KRMS imaging radiometer, as well as Landsat, SSM/I, and ORLS. Other comparison data obtained included video and hand-held photography, voice-recorded observations, and surface measurements from an ice camp in the Beaufort Sea. These SAR data represent a very valuable set of imagery over regionally varied, winter conditions that are particularly useful for verifying the ASF ice classification algorithm.

2. DATA SIMULATION

The data simulation and ice classification flow chart is shown in Figure 1. The aircraft C-band SAR high resolution imagery were used to generate 4-look square-root intensity images of sea ice of sizes 1024 samples by 750 records each (Fig. 2). To simulate the ERS-1 imagery, the 4-look images were convolved with an 8 x 8 smoothing (box) filter which is identical to the procedure used to generate the low resolution (100 m) imagery at ASF. The sample spacing was not changed and kept at 100 m so the resulting images are over sampled. The signal-to-noise (SNR) ratio for the aircraft imagery is about -30 dB which is considerably better than the ERS-1 expected SNR of about -18 dB so a noise level was added to the aircraft data to produce the expected SNR of 9 dB for multiyear ice which has a common backscatter coefficient of -9 dB [1]. Only the near range angles of the aircraft imagery were used (angles 23 to 38 degrees) and a radiometric compensation was then applied to normalize the changes in backscatter due to incidence angle.

3. ICE TYPE BACKSCATTER VARIATION

Preliminary assessment of the aircraft SAR C-band imagery at VV polarization indicates that there is a marked contrast in radar return between multiyear and first year ice. Frazil ice in ice bands and young smooth ice can be distinguished as well as wind-roughened open water leads.

The SAR imagery also detects ridging in both first year and multiyear ice. The separation of these ice types has been strongly confirmed from polarimetric analysis and visual observations as well as from comparison with overlapping KRMS imagery. The KRMS, which has a frequency of 33.6 GHz, is able to distinguish multiyear ice from first year ice and open water but cannot separate new ice from multiyear ice due to similar brightness temperatures [2]

DATA SIMULATION AND ICE-CLASSIFICATION ALGORITHM

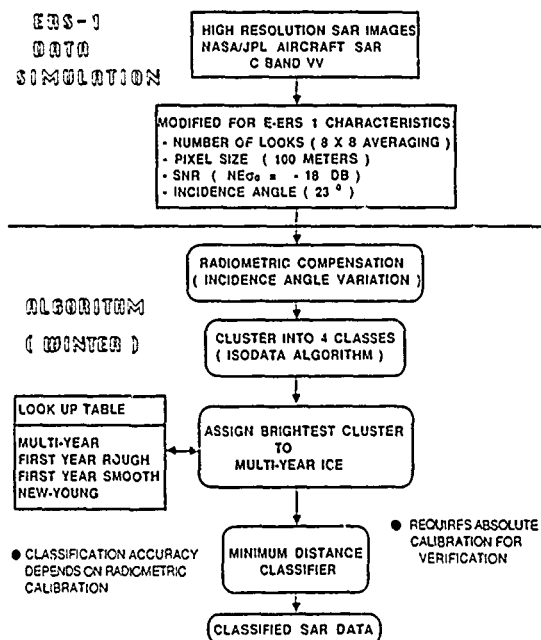


Figure 1. Data simulation and ice classification algorithm flow chart using a lookup table of Arctic winter ice. In this study the lookup table was derived from the aircraft data directly.

The optimum procedure for separating ice types with SAR, upon ice type identification would be to compare calibrated backscatter values to surface-based scatterometer measurements. At present the aircraft C-band SAR has not been highly calibrated. A lookup table of absolute backscatter of various principle ice types obtained with a C-band surface scatterometer is currently being compiled [R. Onstott-personal communication] but was not available in time to be utilized in this preliminary study.

This study determined the relative backscatter power ratios between different ice types from the aircraft SAR imagery itself to serve as a lookup table for the classification algorithm (Fig. 1). Multiyear ice was selected to be the 0 dB reference for comparison with the pixel intensity values of several ice types including the following verified ice types, rough first year ice (greater than 1 m thick), smooth first year ice (less than 1 m thick), young ice, and open water. Figure 3a illustrates the backscatter power of different ice types for one image with varying incidence angles from 23 to 58 degrees after radiometric compensation was performed to remove any backscatter

angular dependency. Four classes are clearly separable in this figure and no open water was present. The ice types are also identifiable on a histogram (Fig. 3b). Next, the backscatter ratio of the different ice types was computed over eight separate images and the results are shown in Table 1. The different classes are separated by about 4 dB. Although there were variations in the multiyear ice intensity values from image to image as high as 4.5 dB, the backscatter ratios between different ice types stayed fairly constant as indicated by the small standard deviation.



Figure 2. NASA DC-8 aircraft SAR imagery from March 11, 1988. Multiyear ice is bright, thick first year ice is medium gray, and thin or young first year ice is darker gray. The incidence angle of the ERS-1 SAR is indicated on the imagery.

4. ICE CLASSIFICATION ALGORITHM

The approach for this classification algorithm is to implement an unsupervised ice segmentation routine which selects a possible set of classes which can then be compared to backscatter values in a lookup table to identify the ice types. The difficulty in long term calibration in SAR systems including ERS-1 precludes having a classification algorithm based on backscatter alone. Multiple lookup tables are needed to account for natural regional and seasonal variations in ice conditions and θ .

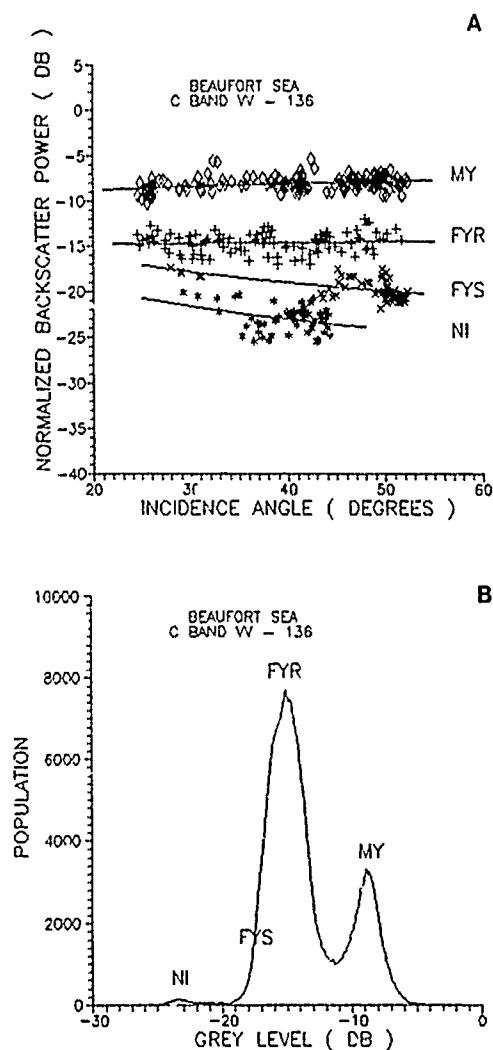


Figure 3. A). Normalized backscatter power of the different ice types from C-band VV aircraft SAR compared to incidence angle after radiometric compensation. B). Histogram of aircraft SAR image.

A clustering algorithm called ISODATA performs the unsupervised segmentation (Fig. 1). Stable clusters can be obtained using a small fraction (3-5%) of the total number of pixels. The input parameters for clustering are simply the desired number of classes and the expected separation between classes in the feature space. For this study four classes separated by at least 4 dB were used although good results can also be obtained with slightly different input parameters. The clusters are ranked according to their mean and then compared to the lookup table. The brightest class is first compared to the multiyear ice table value and then the other classes are compared and defined. Next each pixel in the image is assigned to the nearest class based on the mean grey value computed from a moving 3 x 3 window which also provides some contextual information. This operation corresponds to a minimum distance classifier (Fig. 1).

The clustering algorithm is sensitive to the quality or sharpness of the brightest cluster. Varying the brightest cluster width by 1 dB decreased the classification accuracy 30 %. A typical variation of ± 0.5 dB results in an error of 5-10 % misclassified multiyear ice pixels. This suggests that the ERS-1 long term relative calibration should be accurate to within ± 2.5 dB and short term relative calibration to ± 0.5 dB.

5. RESULTS

Two examples of classified aircraft SAR imagery are shown in Figures 4 and 5. The classification error is estimated to be 7 % for multiyear ice according to the sensitivity studies of cluster sharpness. A comparison of the classified SAR image with KRMS imagery is shown in Figure 6. The comparison of multiyear ice is qualitatively good for both sensors. While not easily seen in this figure, the SAR is able to distinguish young ice from thin first year ice and to detect ridges more clearly than the KRMS.

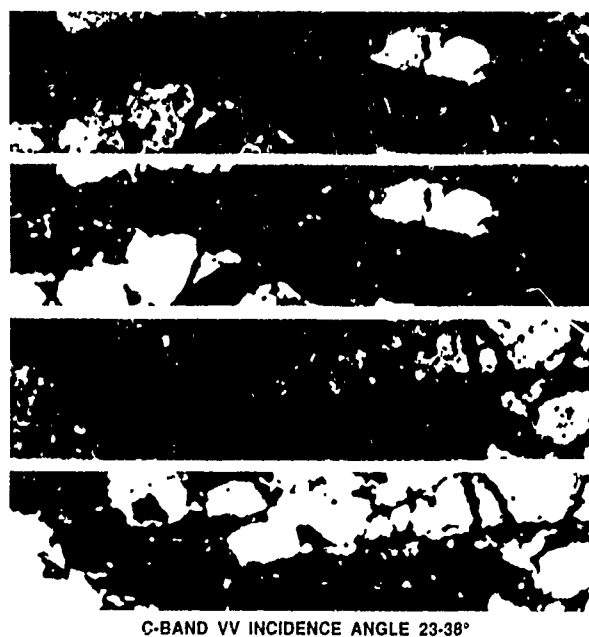
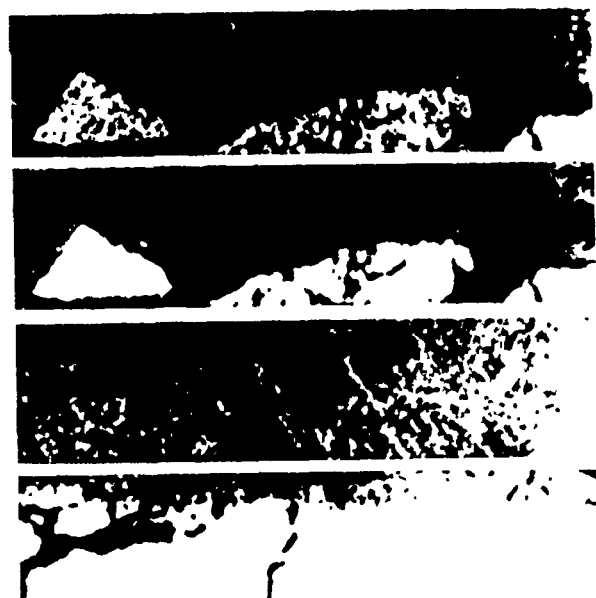


Figure 4. Comparison of aircraft imagery from March 11, 1988 (panels 1 and 3) with classification results (panels 2 and 4).

Further work on this algorithm will include a more comprehensive analysis based on coincident surface ice measurements which will be available in the near future, the investigation of texture analysis in the classification routine, and the incorporation of temperature and wind speed data since these parameters have strong influences on ice and ocean conditions. Also validated lookup tables based on scatterometer data will be included as they become available.

| | NORMALIZED BACKSCATTER POWER RATIO - DC & SAR (DB) | STANDARD DEVIATION OF RATIOS (DB) |
|-------------------|--|---|
| MULTI YEAR | 0 | |
| FIRST YEAR ROUGH | -6.1 | $\pm .8$ |
| FIRST YEAR SMOOTH | -11.6 | $\pm .9$ |
| NEW-YOUNG ICE | -15.4 | $\pm .7$ |

Table 1. Normalized backscatter power ratio of the different ice types using C-band VV data at 25 degree incidence angle.



C-BAND VV INCIDENCE ANGLE 23-38°

RED = NEW ICE
DARK BLUE = FIRST YEAR ROUGH
LIGHT BLUE = FIRST YEAR SMOOTH
WHITE = MULTI-YEAR ICE

Figure 6. Comparison of aircraft imagery (panels 1 and 3) with KRMS imagery (panels 2 and 4) from March 11, 1988.



WHITE = MULTI-YEAR
GREY = FIRST YEAR ROUGH
DARK-GREY = FIRST YEAR SMOOTH
BLACK = NEW ICE

Figure 5. Comparison of aircraft imagery from March 11, 1988 (panels 1 and 3) with classification results (panels 2 and 4).

ACKNOWLEDGEMENTS

The authors wishes to thank D. Eppler and D. Farmer of NORDA at CRREL for the use of the KRMS imagery. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

- [1] Ulaby, F. T., R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive, Vol. III, From Theory to Applications*, Artech House, Dedham, MA, 1986.
- [2] Eppler, D. T., L. D. Farmer, and A. W. Lohanick, Classification of sea ice types with single-band (33.6 GHz) airborne passive microwave imagery, *J. Geophysical Res.*, Vol. 91 (C9), pp. 10661-10695, 1986.

SPECKLE REDUCTION AND MAXIMUM LIKELIHOOD CLASSIFICATION OF SAR IMAGES FROM SEA ICE RECORDED DURING MIZEX 87.

Erik Holbæk-Hanssen (1)
Håkon Tjelmeland (1)
Ola M. Johannessen (2)
Tor Olaussen (2)
Ridvan Karpuz (2)

(1) Norwegian Computing Center, Box 114 Blindern, N-0314 Oslo 3, Norway.

(2) Nansen Remote Sensing Center, Edvard Griegs vei 3A, N-5037 Solheimsvik, Bergen, Norway.

ABSTRACT

The purpose of this study is to develop a system for automatic classification of sea ice from SAR images. A number of speckle reduction filters have been tested. Speckle reduction improves the classification significantly. The adaptive filters the Lee filter and the Sigma filter give the best images for classification. They also preserve details in the images well. When combining backscatter with texture (particularly local variance), small improvements in the classification results are shown. By using supervised classification, we were able to discriminate between three ice categories and open water (possibly mixed with some thin, new ice), with a performance of 92 - 98%.

KEY WORDS: SAR, Sea ice, Speckle, Classification.

1. Introduction

Interpretation of sea ice in the Arctic regions will be an important application of the SAR data produced by the ERS-1 remote sensing satellite. Real time production of reliable sea ice data will become crucial for future Arctic activities. This study aims at providing fast and reliable systems for interpretation of SAR images of sea ice. The study is supported by the Royal Norwegian Council for Industrial Research (NTNF).

Similar studies are reported in a number of publications. See for instance (Holmes, 1984) and (Skriver, 1986).

This study is based on the in situ data recorded during the MIZEX 87 experiment. The in situ data are described in Section 2.

The first results from classification of the SAR images gave poor results, mostly due to the speckle noise in the images. A number of methods for speckle reduction was therefore implemented. The results of these tests are described in Section 3.

In order to improve the classification results further, local texture in the SAR images was computed. The results from these tests are described in Section 4.

In Section 5 is described classification of the images using maximum likelihood classification. First, classification into the seven ice types defined from the in situ data was tried. The experiments proved that a reduction into four categories was optimal.

2. In Situ Data

A number of experiments have been performed in order to study the behavior of SAR images. Some studies are based on SEASAT or SIR-A or SIR-B satellite data. Other studies are based on airborne sensors. An important aspect of the experiments is to provide reliable surface truth of the areas covered by the satellites. The in situ data may be used for calibration of the planned methods for automatic interpretation of the data achieved from the SAR sensors.

The data used in this study was acquired during the MIZEX 87 experiment that took place in April 1987 in the Fram Strait, see (Johannessen, 1988) and (Shuchman, 1988).

22 SAR data collection missions were flown daily, using the Intera STAR systems. The SAR data recorded are 7-look X-band, HH-polarized, and the resolution is 15 x 15 m. This study is based upon a scene consisting of 3000 x 2000 pixels.

During the experiment, the surface truth was recorded by a research vessel and helicopters. 32 areas within the image were categorized into the following seven categories (after Shuchman, 1988):

- A. 50 - 60% Multiyear Ice in Consolidated First-Year Ice Framework (1-3 cm snow).
- B. 30 - 40% Multiyear Ice with Rubble in Loose First Year Ice Framework (some open water).
- C. First-Year Ice with Rubble (.60-1.5 m thick).
- D. First-Year Ice (20-40 cm thick).
- E. New Ice (5-8 cm thick).
- F. Open Water.
- G. Open Water with Grease Ice Streamers.

The classification tests in this study are based on the in situ data from these (rather small) areas.

3. Speckle reduction

Speckle noise is disturbing for the interpretation of SAR images. Much emphasis has therefore been put in reduction of the speckle, see for instance (Skriver, 1986), while maintaining as much as possible of the information in the image, such as texture and edges. Most important in our choice of filters is to find those which give the best images for classification.

We have in our study applied a number of methods for

noise reduction. These include traditional image processing methods as average and median calculations, as well as lowpass filters. Because the speckle noise is data dependent, filters that can adapt to the image is likely to give the best results. Various adaptive filters have been implemented and tested on the SAR imagery. These include the filters called "Lee" (Lee, 1980), "Frost" (Frost, 1982), "Sigma" (Lee, 1983), "Symmetric nearest neighbour" (Harwood, 1984) and "Maximum Homogeneity" (Nagao, 1978) filters. Among these adaptive filters, we found that the Lee and the Sigma filters gave the best results, and they were used for further tests.

For comparison, five filters, which have approximately equal smoothing capabilities, were applied to the same image, shown in fig. 1. (They give approximately the same results when applied to an image without any other variations than the speckle noise.) The five filters which have been tested are: Average (with a Gaussian shaped kernel of 9x9 pixels), Median with kernel 7x7, a lowpass filter, the Sigma filter kernel 5x5 (used twice), and the Lee filter with kernel 7x7.

In table 2 in Section 5, some classification results are summarized. We see that speckle-reduced SAR images give much better classification results than unfiltered images. Up to a certain limit, we have found that the better the images are smoothed, the better are the classification results. Our experience is that filter kernels of 7x7 or 9x9 pixels do well.

To compare the filters' ability to preserve structures, the backscatter values along a fixed line (shown on original) were plotted as "profiles" for the same images, see fig. 1.

From fig. 1, we can see that the adaptive filters blur the images less than the non adaptive filters. The median-filter also preserve much of the structures well, but the highest peak has almost disappeared.

From these studies and the classification results shown in section 5, we conclude that the most successful filters are the two adaptive filters Lee and Sigma.

4. Texture

One may expect that the classification will be improved by including textural information in discriminating between the classes in SAR images. See for instance (Holmes, 1984), (Skriver, 1986). In our study, we have applied different textural measures, including: variance and standard deviation, power-to-mean, local energy and local frequency distribution.

Some results of the classification are shown in table 2. Only marginal improvements were obtained by adding texture: combined with the average, lowpass or Lee filters, the error rates were reduced, but combined with the median or the Sigma filters, no improvement was obtained at all. Variance, standard deviation and power-to-mean give the best results. Note, however, that in our experiment the classification results were already very good when using backscatter only.

We also found that texture alone is not sufficient for discriminating between these ice categories. For discrimination between only ice and open water, however, variance or standard deviation succeed with more than 95% correct classification.

5. Data Analysis and Classification

The main purpose of this study is to determine how well the 7 data classes described above classes can be separated, or which classes can be separated. We did this by first studying the data by histograms and statistical measures, and then classifying the data using maximum likelihood classification.

The backscatter values and the textural measures were calculated for the test areas described above, where the ice types are known. Histograms for each of the classes were computed. Fig 2 shows the histograms for the 7 ice categories for unfiltered and filtered images.

The relationships between the classes were studied by computing the Mahalanobis distances, which are statistical measures describing the distances between the classes. Table 1 shows the Mahalanobis distances between the ice categories. Fig. 3 shows the distribution of the categories when backscatter and texture are combined.

| | A | B | C | D | E | F | G |
|---|-------|-------|------|------|------|------|---|
| A | 0 | | | | | | |
| B | 2.86 | 0 | | | | | |
| C | 3.42 | 2.24 | 0 | | | | |
| D | 9.22 | 6.32 | 2.37 | 0 | | | |
| E | 13.84 | 11.02 | 4.61 | 6.94 | 0 | | |
| F | 13.85 | 10.98 | 4.15 | 7.25 | 2.24 | 0 | |
| G | 14.64 | 11.90 | 5.77 | 8.37 | 2.85 | 3.70 | 0 |

Table 1. Mahalanobis distances between 7 sea ice and water categories. Backscatter (Lee-filtered) combined with texture (Variance). As a rule, we can state that when the distance between two classes is greater than 4.0, the classes may be well separated, when it is less than 4, confusion between the classes is more likely to occur.

These studies indicate that we cannot distinguish between all the 7 classes. We wanted therefore to categorize the ice into fewer categories. From the data analysis we found that:

- Type A may be kept as a separate ice category.
- Type C can be described as a mixture between the two classes B and D, and the two mixed ice categories B+C1 and C2+D may be distinguished as two separate ice categories.
- Types E, F and G can hardly be separated from each other, but are well separated from the other classes.

Discrimination between these 4 classes seems therefore to be achievable: A, B+C1 C2+D and E+F+G.

The images were classified by using supervised classification. In order to get realistic estimated of the classification results, the in situ data set was divided into two disjoint groups. The first data set was used as a training set when calculating the probability density functions used in the classification. When classifying, confusion matrices and performance statistics was calculated for the second data set. The results show an overall performance of maximum 60 - 70% when using all the 7 classes. Multiyear Ice in Consolidated First-Year Ice Framework (ice category A) can be fairly well discriminated from the others (about 90%), but the other classes are too intermixed with each other.

When rearranging the class definitions into the four classes described above, we obtained the classification rates described in table 2.

| Filter | Original | Average | Median | Low pass | Sigma | Lee |
|------------------|----------|---------|--------|----------|-------|-----|
| Backscatter only | 74 | 95 | 97 | 95 | 98 | 96 |
| With texture | 78 | 98 | 97 | 98 | 98 | 98 |

Table 2. Percent correctly classified pixels for unfiltered and filtered images.

Fig. 4 shows results from classifying a SAR image into 4 classes.

When instead using data set 2 for training and testing on data set 1, we obtained similar results, but the performance was reduced to 92 - 94 %. There, texture only improves the results by 1 - 2 %.

The results of these tests show that when reliable in situ data are available, one is able to classify into four sea ice and water categories with more than 90% significance.

When limiting ourselves to only two classes: (1) Sea Ice and (2) Open water and new ice, the classification results were almost 100% correct.

6. Future plans.

The objective of this project is to make a system that can produce the results in real time. Filter processing and texture analysis are CPU-consuming tasks. We have therefore applied an image processing system with a fast filter processor for the computations (GOP-300). Typical processing times for a 512x512 image are 1 - 2 minutes for filtering and texture analysis, and similar performance for classification. The plans are to integrate the filtering and the classification into one process.

The other plans for the project is to combine the supervised classification with unsupervised classification, in order to classify sea ice in the more realistic situation, where no, or very limited in situ data are available.

7. Conclusion

We have shown that speckle reduction is necessary for classification of SAR images of sea ice. The adaptive filters Lee filter and Sigma filter give the best results, and they also preserve details in the images well.

When combining backscatter with texture (particularly variance), small improvements in the classification results are shown.

By using supervised classification, we were able to discriminate between three ice categories and open water, with a performance of 92 - 98%. We were able to discriminate between ice and open water (possibly intermixed with new, thin ice) with almost 100% accuracy. These results are based upon reliable in situ data, but the test areas are rather small.

References

- V.S. Frost et.al.: "A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise." IEEE Transactions on PAMI, Vol. PAMI-4, no. 2, March 1982.
- J.W. Goodman: "Some Fundamental Properties of Speckle," Journal of Opt. Soc. Am., Vol. 66, No 11, November 1976.
- D. Harwood et.al.: "A New Class of Edge-Preserving Smoothing Filters." Report CAR-TR-59, University of Maryland, 1984.
- E. Holbæk-Hanssen et.al.: "Filtering and Classification of SAR Images from Sea Ice." Report No. 821, Norwegian Computing Center, Oslo, 1989 (in Norwegian).
- Q.A. Holmes et.al.: "Textural Analysis and Real-time Classification of Sea-ice Types Using Digital SAR Data." IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-22, No. 2 March 1984.
- O.M. Johannessen et.al.: "Overview of the Winter Marginal Ice Zone Experiment in the Greenland and Barents Seas." Proceedings of the Seventh International Conference on Offshore Mechanics and Arctic Engineering, Vol. IV, pp. 99-109, 1988.
- J-S Lee: "Digital Image Enhancement and Noise Filtering by Use of Local Statistics." IEEE Transactions on PAMI, Vol. PAMI-2, no. 2, March 1980.
- M. Nagao et.al.: "Edge Preserving Smoothing." Computer Graphics Image Processing 9, 1979.
- R.A. Shuchman et.al.: "Intercomparison of Synthetic and Real-Aperture Radar Observations of Arctic Sea Ice During Winter MIZEX '87." In Proceedings from IGARSS '88, Edinburgh, Sept. 1988.
- H. Skriver et.al.: "Active Microwave Observations of Sea Ice and Icebergs." In "SAR Application Workshop", Proceedings of an ESA Workshop held at Frascati, Italy, Sept. 1986.

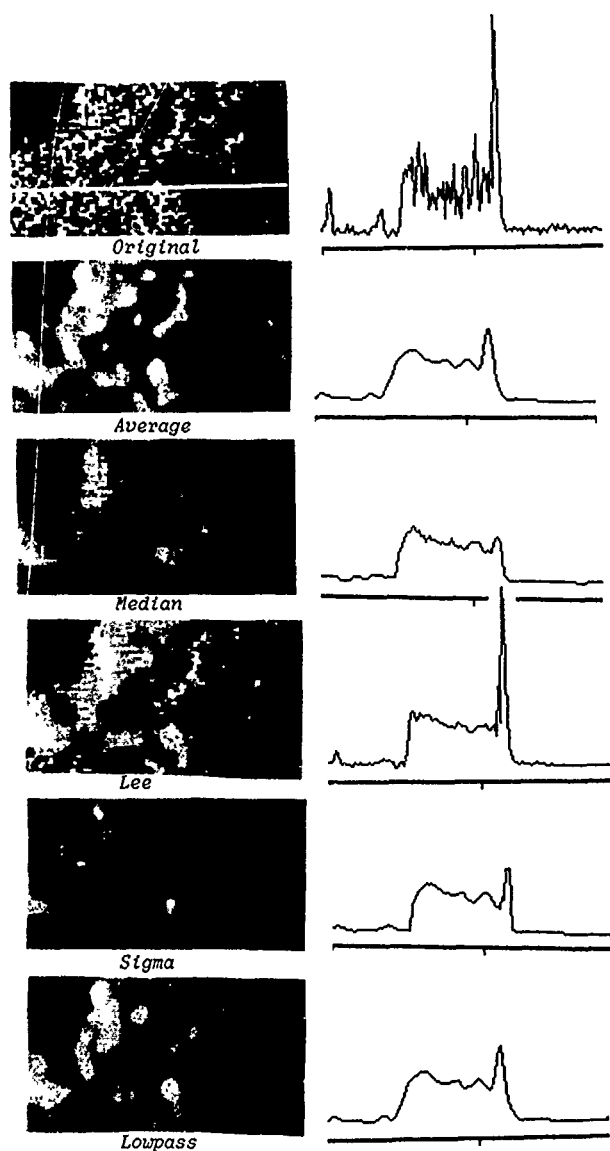


Fig. 1. SAR image filtered by different filters having approximately the same smoothing capabilities on a "smooth" image. To the right are shown backscatter values along a fixed line in the images.

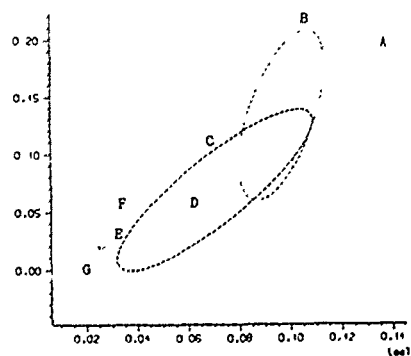


Fig. 3. Idealized scattergrams of ice categories, combining backscatter (Lee-filtered) and the Power-to-mean Ratio.

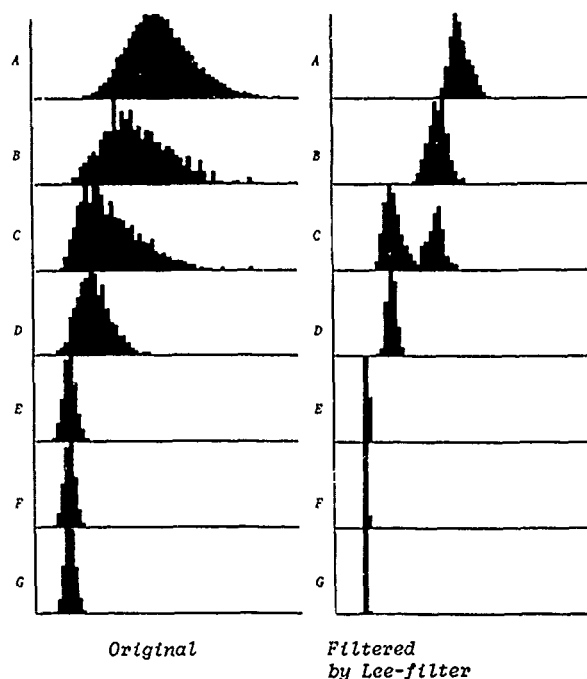


Fig. 2. Histograms of ice categories taken from the in situ data.

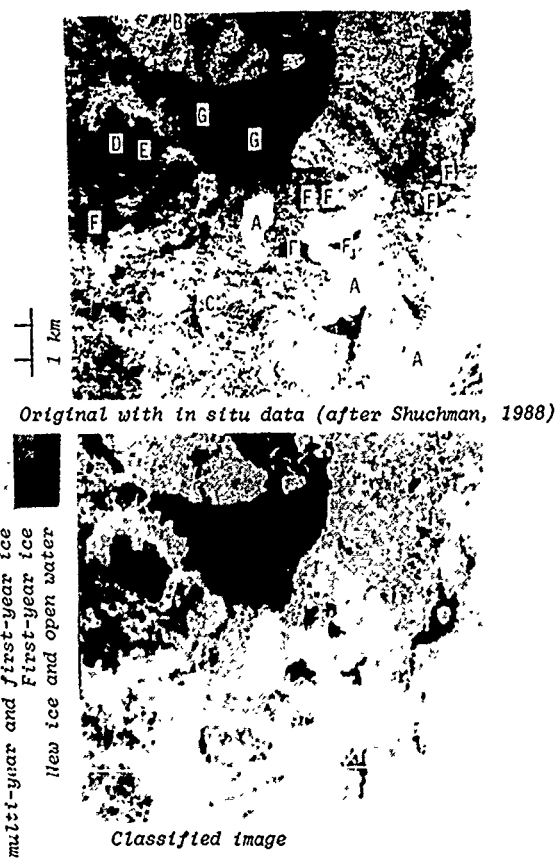


Fig. 4. Result from classification into four ice categories.

MULTIVARIATE ANALYSIS OF TEXTURE STATISTICS FOR SAR SEA ICE DISCRIMINATION

David G. Barber and Ellsworth F. LeDrew

Earth-Observations Laboratory, Institute for Space and Terrestrial Science
Department of Geography, University of Waterloo
Waterloo, Ont. N2L 3G1

Abstract

In this paper we report on discrimination of sea ice classes using texture statistics derived from Grey Level Co-occurrence Matrices (GLCM). Simple Discriminant Analysis and Multivariate Discriminant Analysis are used to interpret the relationship between 5 texture measures, GLCM inter-pixel sample distances (δ) and orientation (α) on discrimination of 4 ice classes. Results show that discrimination is maximized when measures at different α and δ are used. Discriminability is affected by adaptive filtering and resampling to a coarser pixel scale.

Keywords: Discriminant analysis, Texture statistics, Synthetic Aperture Radar, Sea ice.

1. Introduction

Quantitative description of radar backscatter is central to feature discrimination in radar remote sensing applications research. This project is directed towards description of *functional* ice classes, derived from Grey Level Co-occurrence Matrix (GLCM) probability distributions or more specifically, from texture statistics derived from these matrices. The GLCMs used consist of the frequency of joint occurrences of quantized grey levels (4 bit), at intersample spacing distances ($\delta=1,3,5$) and orientations ($\alpha=0^\circ, 45^\circ, 90^\circ$). Five texture measures, commonly found in the literature, are used in this investigation [2] TO [6] (Haralick, 1986). Our intention is to determine the utility of these texture measures for discriminating ice classes. Four objectives are addressed:

- What is the relative contribution of each texture measure to discrimination of ice classes?
- What affect does the intersample spacing variable δ and orientation variable α have on discrimination of ice types.
- Can discriminant analysis be used to indicate the minimum number of texture measures required to achieve i) a reasonably precise discrimination ii) complete discrimination.
- What affect does filtering, and resampling to 102m pixel size, have on discrimination of ice classes.

2. Methods

A STAR-1 SAR image (Mould Bay, East-West pass, May 13, 1984) from the Canadian Radar Age/Type Algorithm Group (CRAGTAG) standardized image dataset, was used to create 3 images: original 6m (7 look), 6m adaptively filtered (6mf), and

102m resampled image (CRAGTAG Tape #RD2826). The 6mf image was produced using an adaptive filter [1] described in Durand *et al.* (1987) where the subscript 'nk' is a 3 by 3 window and 'hom' is a homogeneous texture class.

$$\hat{X}_{ij} = \bar{X}_{nk} + (X_{ij} - \bar{X}_{nk}) \cdot (s_{nk}^2 - s_{hom}^2) / s_{nk}^2 \quad [1]$$

All ice classes were selected using a 20 by 20 pixel grid, overlain on a hardcopy image. Ground confirmation was supplied by the Atmospheric Environment Service Ice Branch, as part of the standard CRAGTAG dataset. Random coordinates were selected using a Macintosh II pseudo random number generator. A subarea was accepted if it represented a homogeneous sample of an ice type of interest. This process was repeated until 5 homogeneous replicates of 4 ice types (New Ice (NWI), First-Year Ice (FYI), Multi-Year Ice (MYI) and Land (LAND)) were obtained. The class, New Ice, did not constitute a sufficient number of pixels to select 5 independent samples at the 102m pixel size. Only FYI, MYI and LAND are used in the discriminant analysis of 102 m data.

The Grey Level Co-occurrence Matrices and the texture statistics derived from these matrices were programmed on a Macintosh II microcomputer, based on algorithms described in Haralick (1986) and Shanmugan *et al.* (1981).

$$\text{Uniformity} \quad \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2 \quad [2]$$

$$\text{Correlation} \quad \sum_{i=1}^n \sum_{j=1}^n \frac{(i - \mu_x)(j - \mu_y) C_{ij}}{\sigma_x \sigma_y} \quad [3]$$

$$\text{Entropy} \quad - \sum_{i=1}^n \sum_{j=1}^n C_{ij} \log C_{ij} \quad [4]$$

$$\text{Dissimilarity} \quad \sum_{i=1}^n \sum_{j=1}^n C_{ij} |i - j| \quad [5]$$

$$\text{Contrast} \quad \sum_{i=1}^n \sum_{j=1}^n C_{ij} (i - j)^2 \quad [6]$$

Where: C_{ij} is the matrix cell frequencies normalized by the number of pixel pairs P_{ij} .
 n is the number of pixel pairs in the image at (α, δ) .
 σ_x is the sample standard deviation of row i .
 σ_y is the sample standard deviation of column j .
 μ_x is the sample mean of row i .
 μ_y is the sample mean of column j .

To assess the utility of texture statistics for discrimination of ice classes in SAR imagery two forms of linear discriminant analysis were conducted.

1) Simple Discriminant Analysis (SDA) consists of a pairwise linear combination of variables to maximize separation of 2 groups. The SDA was programmed on a Macintosh II microcomputer, based on algorithms described in Ludwig *et al.* (1988). Output from SDA provides a pair of discriminant coefficients, the relative percent contribution of each variable to discrimination, and a Mahalanobis multivariate distance measure (D^2); which is the metric used for class assignment. All pairwise combinations of variables (5 texture statistics at $\alpha=0^\circ$ and $\delta=1$) and classes (4 ice types) were tested using SDA.

2) Multiple Discriminant Analysis (MDA) is a statistical technique which can be used to define spatial patterns and to assist in meaningful interpretation of these patterns (Williams, 1983). We used MDA as a means of determining which set of p variables maximize discrimination of k ice classes. Identification of classes is done *a priori*, either through functional description of classes (as we have done), or through some form of ordination. The number of discriminant functions that result is the smaller of $k-1$ or p . MDA finds a set of linear transformations which maximize the inter- to intra-class variation over k . This is equivalent to maximizing the F ratio of a one way analysis of variance (Manly, 1986). The first function (Z_1) maximizes the inter- versus intra-class variation, over the entire data matrix. The second function (Z_2) maximizes the F ratio, with the important limitation that Z_2 is orthogonal to Z_1 , and so on (Manly, 1986).

Two approaches to discriminant analysis are available for interpretation of feature discrimination: predictive and descriptive (Johnston, 1980). In this analysis, descriptive MDA is used to gain an understanding of the complex relationships between groups of texture statistics and ice class discrimination. By transforming the original texture variables through each discriminant function, we can display relationships between the ice classes in discriminant space (Fig. 2). Interpretation of these plots provides information on the direction and magnitude of class separation as a function of the input variables. Other, more objective, measures used in interpretation of MDA include Chi-Square statistics for the eigenvalues associated with each discriminant function, and an overall appropriateness test for a specific set of variables (Wilks Lambda). Mathematical calculation of these statistics and a description of hypothesis conditions is provided in Manly (1986).

Assumptions of these parametric analysis procedures are that the variance-covariance matrices are samples of the same population matrix, and that within each class the variables follow a multivariate normal distribution (Manly, 1986). Because of correlations within each statistic over the 3 levels of α and δ , and similarities in calculation of the statistics (eg. Dissimilarity [5] and Contrast [6]), we expect a high degree of multicollinearity. To decrease this problem a square root transformation was applied to the data matrix. Although discriminant analysis is considered invariant to scale (Manly, 1986) a row column centering operation appeared to improve discrimination. The technique of centering data matrices, in discriminant analysis, is described in Legendre and Legendre (1983).

Multiple Discriminant Analysis (MDA) was conducted on 7 data sets (Set #); (Vars) is the number of texture statistics used in

each trial; (α) and (δ) are orientation and interpixel sampling distances used to generate the GLCM; (Trials) is the number of sets of variables (ie; Set #1=3 trials of 5 vars.); (Selection) is either fixed (all possible) or random (Table 1). Set # is referenced in the results.

Table 1. Parameters contained in each MDA dataset.

| Set # | Vars | α | δ | Trials | Selection |
|-------|------|------------|----------|--------|-----------|
| 1 | 5 | 00 | 1,3,5 | 3 | fixed |
| 2 | 5 | 00,450,900 | 1 | 3 | fixed |
| 3 | 3 | 00 | 1,3,5 | 15 | random |
| 4 | 5 | 00 | 1,3,5 | 20 | random |
| 5 | 10 | 00 | 1,3,5 | 3 | fixed |
| 6 | 10 | 00 | 1,3,5 | 20 | random |
| 7 | 10 | 00,450,900 | 1 | 3 | fixed |

3. Results and Discussion

First-order image statistics for each subarea are presented to provide the reader with an indication of inter- to intra-class variability over the ice classes tested. Note that mean and standard deviation (SD) are the average of 5 replicates. The SD decreases from the original 6m data to filtered 6m and resampled 102m data. The substantial drop in SD for 102m data is a function of resampling, and contributes to difficulties in discriminating ice classes using texture statistics.

Table 2. First order image statistics for each texture field.

| Class | 6m data | | 6m filtered | | 102m data | |
|-------|---------|-------|-------------|-------|-----------|------|
| | Mean | SD | Mean | SD | Mean | SD |
| FYI | 27.35 | 8.37 | 27.26 | 6.31 | 27.04 | 3.75 |
| LAND | 54.93 | 20.45 | 55.62 | 18.46 | 47.57 | 6.80 |
| MYI | 44.32 | 15.54 | 44.63 | 13.52 | 43.00 | 5.18 |
| NWI | 19.34 | 5.33 | 19.37 | 3.59 | | |

Simple Discriminant Analysis (SDA)

To assess discriminability of ice classes, using the 5 texture statistics, a Simple Discriminant Analysis (SDA) was conducted. The Mahalanobis distance measure provides an index of class separability, relative to pairs of texture statistics at $\alpha=0^\circ$ and $\delta=1$. The results are summarized by the magnitude of D^2 (larger D^2 = more separation), and by whether the separation of group centroids is significant. Differences due to filtering (6mf) and resampling (102m) are evident by comparing the D^2 statistics (Table 3). Comparison of D^2 distances should be considered an index of relative difference. With such low degrees of freedom, the statistic may not be stable (Manly, 1986).

Discrimination of Multi-Year Ice and Land was the most difficult (Table 3). In 6m data these ice types can only be differentiated using a single pair of texture statistics; Entropy and Dissimilarity (6.45 in Table 3). The next nearest value, Uniformity-Dissimilarity (2.83), illustrates how different this pair of measures is from the rest. All other classes are easily discriminated in the 6m data. After filtering (6mf), 4 pairs of variables could discriminate between the difficult classes; MYI vs LAND. Again, Entropy-Dissimilarity provides the best discrimination. Entropy-Contrast provides a strong second with the pairs Uniformity-Dissimilarity and Uniformity-Contrast also providing significant discrimination. As a general result, filtered data provide the most powerful discrimination, followed by original 6m data, then the resampled imagery. Note that no pair of texture measures were able to discriminate MYI-LAND in the 102m image. This is most likely due to a reduction in the contrast between texture classes in the 102m image. Removing the component of variation due to fading in the 6mf imagery, contributes to improved discrimination.

Table 3. Summary of Simple Discriminant Analysis (SDA) conducted on all pairwise comparisons of 5 texture measures on 4 ice classes. Mahalanobis multivariate distances are presented. F statistic significance is based on calculation in Ludwig (1988).

| Ice Class | Image | Unif-Corr | Unif-Ent | Unif-Diss | Unif-Cont | Corr-Ent |
|-------------|-------|-----------|----------|-----------|-----------|----------|
| MYI vs LAND | 6m | 1.06* | 0.89* | 2.83* | 1.37* | 1.03* |
| | 6mf | 1.52* | 1.58* | 5.30 | 4.88 | 1.31* |
| | 102m | 0.69* | 1.00* | 0.41* | 0.42* | 0.35* |
| MYI vs FYI | 6m | 11.33 | 11.90 | 13.47 | 11.33 | 12.89 |
| | 6mf | 10.70 | 11.91 | 12.68 | 10.66 | 21.76 |
| | 102m | 20.63 | 2.59* | 9.30 | 8.71 | 20.61 |
| LAND vs FYI | 6m | 14.25 | 13.30 | 13.31 | 13.51 | 16.12 |
| | 6mf | 23.11 | 20.11 | 20.22 | 21.13 | 32.64 |
| | 102m | 64.91 | 10.21 | 9.12 | 11.04 | 64.70 |
| MYI vs NWI | 6m | 18.02 | 33.73 | 25.09 | 17.47 | 30.59 |
| | 6mf | 26.89 | 28.16 | 31.55 | 26.62 | 29.61 |
| | 102m | | | | | |
| LAND vs NWI | 6m | 15.68 | 24.79 | 19.66 | 18.25 | 34.58 |
| | 6mf | 30.57 | 33.67 | 32.64 | 31.42 | 43.46 |
| | 102m | | | | | |
| FYI vs NWI | 6m | 10.02 | 8.31 | 11.23 | 10.50 | 8.73 |
| | 6mf | 19.78 | 14.54 | 29.99 | 38.46 | 16.30 |
| | 102m | | | | | |

| Ice Class | SDA | Corr-Diss | Corr-Cont | Ent-Diss | Ent-cont | Diss Cont |
|-------------|------|-----------|-----------|----------|----------|-----------|
| MYI vs LAND | 6m | 1.17* | 1.08* | 6.45 | 2.84* | 0.81* |
| | 6mf | 1.22* | 1.24* | 12.92 | 11.68 | 0.45* |
| | 102m | 0.57* | 0.40* | 0.84* | 0.29* | 1.72* |
| MYI vs FYI | 6m | 25.25 | 5.26 | 12.63 | 12.78 | 38.16 |
| | 6mf | 29.28 | 4.68 | 11.41 | 17.00 | 53.94 |
| | 102m | 22.24 | 21.91 | 9.40 | 9.56 | 5.99 |
| LAND vs FYI | 6m | 12.17 | 10.05 | 11.75 | 13.89 | 14.85 |
| | 6mf | 16.08 | 10.31 | 22.72 | 34.63 | 29.92 |
| | 102m | 71.18 | 70.96 | 4.63 | 6.05 | 7.14 |
| MYI vs NWI | 6m | 29.65 | 10.64 | 30.19 | 31.78 | 33.83 |
| | 6mf | 50.81 | 9.81 | 27.42 | 30.68 | 69.59 |
| | 102m | | | | | |
| LAND vs NWI | 6m | 25.00 | 17.79 | 35.54 | 32.37 | 24.44 |
| | 6mf | 35.42 | 18.21 | 36.49 | 40.78 | 53.15 |
| | 102m | | | | | |
| FYI vs NWI | 6m | 11.01 | 10.54 | 10.26 | 12.36 | 10.53 |
| | 6mf | 27.01 | 26.92 | 24.82 | 34.22 | 26.13 |
| | 102m | | | | | |

* Denotes F statistics which are not significant at F 0.10; (2, N1+N2-3) df.

— Denotes a flag for references in the text.

Multiple Discriminant Analysis (MDA).

Results of SDA are useful in identifying MYI-LAND as the pair of ice classes most difficult to differentiate. We used MDA to conduct exploratory data analysis on the location and separation of ice classes in discriminant space to try and determine the minimum number of variables needed to achieve i) minimal error, and ii) no errors in discrimination.

MDA was applied to discrimination of 4 ice types: Multi-Year Ice, First-Year Ice, New Ice and Land, using 6m data. To identify differences that might exist in discrimination due to α and δ a set of MDA (Table 1) were run on all five texture statistics from Set 1 (Fig. 1a-c) and Set 2 (Fig. 1d-f). The discrimination tables (Fig. 1a-f) show the observed versus predicted group discrimination based on a multivariate D^2 distance metric. An intersample pixel spacing of $\delta=1$ provides the least discrimination error (7 percent). At $\delta=3$ and 5, errors occurred between classes of similar texture (LAND vs MYI and FYI vs NWI). It is interesting to note the error in discrimination that occurs at $\alpha=90^\circ$. One observation of MYI was misclassified as FYI, even though the first-order statistics for these two classes are considerably different (Table 2). Discriminant plots of this analysis show that the misclassification was both a function of increased dispersion of the FYI class and a slight increase in overlap between the

clusters of points for MYI and FYI, compared with 0° and 45° orientations.

When sets of three variables were randomly selected from the 15 available (Set 3, Table 1), errors in discrimination ranged from 15 to 40 percent. We found the minimum number of variables needed to obtain only a single error in discrimination was 5. Using each of the 5 texture measures at $\delta=1$ and $\alpha=0^\circ$ resulted in a single error between NWI and FYI (Fig. 1). When Set 4 was submitted to a MDA, four other combinations of texture measures showed the same error rate. Uniformity at $\delta=1$ and $\delta=5$ with Entropy at $\delta=1$ and $\delta=5$ and one of either correlation or contrast at either $\delta=1$ or $\delta=5$, results in a single error in discrimination between LAND and MYI. The other 16 sets of variables (Set 4) resulted in errors from 13 to 27 percent. Apparently, Uniformity and Entropy have a significant impact on discrimination of the ice classes tested. Perhaps more importantly, the combination of $\delta=1$ and $\delta=5$ for these measures, is important in discrimination.

Similar exploratory analysis was conducted to determine the minimum number of variables needed to obtain complete separation of ice classes. From Set 5 only 1 combination of variables provided error free discrimination; each of the 5 texture statistics at $\delta=1$ and $\delta=5$. The other combinations of δ (1-3 and 3-5) did not provide error free discrimination. To determine the uniqueness of this result 20 other combinations of 10 variables (Set 6) were randomly selected and submitted to a MDA. Error of discrimination in these trials ranged from 10 to 25 percent. Although using 10 texture measures is far from operationally viable for SAR sea ice discrimination, this result provides another example of how contrasts in the parameter δ appear to be useful in discrimination of ice classes.

Interpretation of MDA for the effect of orientation (Set 7) shows the combination $\alpha=0^\circ$ and 90° provides full discrimination (Fig. 2a). The other two combinations $\alpha=45^\circ$ and 90° (Fig. 2b) and $\alpha=0^\circ$ and 45° (Fig. 2c) provide full discrimination and 1 error, respectively. The Chi-Square statistics measure the significance of each discriminant function (cumulative measures in Table 4) to separation of ice classes. Although $45^\circ-90^\circ$ is error free, the discrimination is not as powerful as $0^\circ-90^\circ$ (Table 4). Figure 2a-c is very similar to discriminant plots of Set 5. At high contrast (either δ at 1 and 5, or α at 0° and 90°), a more powerful discrimination is observed. Decrease in the Chi-Square statistic over these levels of contrast are consistent when both α and δ are considered (Table 4). Note that orientation appears to provide stronger discrimination than the equivalent variables, which contrast interpixel spacing (Table 4).

Table 4. Relationship between contrast in orientation and interpixel sampling distance for all 5 texture measures. Chi-square statistic calculation is from Manly (1986).

| Chi-Square | GLCM orientation | | | Interpixel spacing | | |
|-------------|------------------|-------|------|--------------------|-------|-------|
| | 0-90 | 45-90 | 0-45 | 1 & 5 | 3 & 5 | 1 & 3 |
| 1 through 3 | 73.2 | 68.9 | 53.6 | 89.2 | 83.8 | 63.9 |
| 2 through 3 | 37.8 | 40.2 | 26.0 | 35.1 | 37.4 | 33.8 |
| 3 through 3 | 11.6 | 17.6 | 10.2 | 10.8 | 12.3 | 11.3 |

| | FYI | NWI | LAND | MYI | FYI | NWI | LAND | MYI | FYI | NWI | LAND | MYI |
|------|-----|-----|------|-----|-----|-----|------|-----|-----|-----|------|-----|
| FYI | 5 | | | | 5 | | | | 5 | | | |
| NWI | 1 | 4 | | | 1 | 4 | | | 1 | 4 | | |
| LAND | | | 5 | | | | 4 | 1 | | | 4 | 1 |
| MYI | a | | | 5 | b | | 1 | 4 | c | | | 5 |
| FYI | 5 | | | | 5 | | | | 5 | | | |
| NWI | 1 | 4 | | | 1 | 4 | | | 1 | 4 | | |
| LAND | | | 5 | | | | 4 | 1 | | | 4 | 1 |
| MYI | d | | | 5 | e | | | 5 | f | 1 | | 4 |

Figure 1 Discrimination errors from 5 texture statistics, Set 1 (a, b, c) and Set 2 (d, e, f), using 106m data

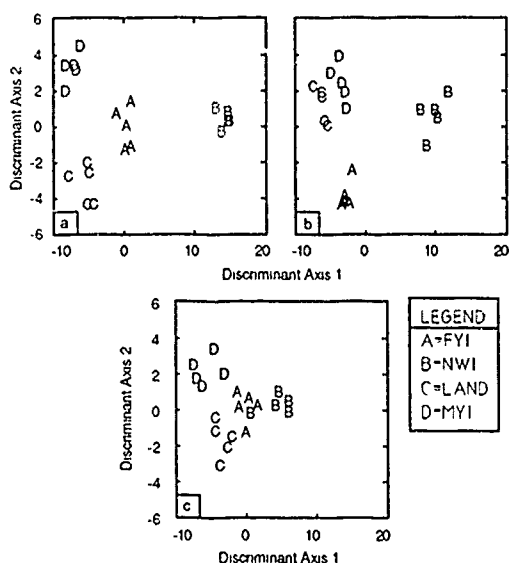


Figure 2 Plot of the two significant discriminant functions for 10 texture statistics (a=0-90 deg, b=45-90 deg and c=0-45 deg)

Effect of filtering and resampling on discrimination

Theoretically, adaptive filtering will remove a component of the variation attributable to fading, leaving a texture random variable which is more representative of the natural variation of ice type backscatter (Ulaby, *et al.* 1986). Simple Discriminant Analysis showed that, in general, filtering improved discrimination. In this section we wish to determine if filtering or resampling changes the discriminatory power of the sets of variables tested on the 6m data (Table 1). The Chi-Square statistic, associated with the discriminatory power of Z_1 to Z_3 is compared between the three image types. Although the sample sizes are much too small for definitive statements, the data does indicate a similar relationship determined with SDA.

When all 5 variables were used at one of the intersample spacings ($\delta=1, 3$, and 5) the results show that filtered images consistently resulted in a more powerful discrimination than 6m or 102m images. Groups of 10 variables consistently showed improved discrimination in the filtered imagery, but not in the resampled imagery (102m). Groups of 3 texture measures showed no difference in discrimination on either the filtered or resampled imagery. We make the general conclusion that with a small number of variables, changes in image texture caused by filtering or resampling are not captured by the discriminant

functions. With 5 or more discriminant variables the decrease in fading caused by filtering is reflected in improved discrimination of ice classes. The decrease in contrast between ice texture when the imagery is resampled to 102m results in poorer discrimination only when the number of variables exceeds 5.

4. Conclusions

Our analysis indicates that the pair Entropy-Dissimilarity contribute significantly to discrimination of MYI and LAND in the 6m image. Correlation provides poor discrimination, paired with one of the other measures. The parameters δ and α appear to be useful when statistics are obtained at contrasting levels (i.e., $\delta=1$ and 5 , $\alpha=0^\circ$ and 90°) for the same statistic. Using all 5 measures provides a reasonable discrimination, while 10 measures (all 5 at contrasting α or δ) are needed for complete discrimination. Adaptive filtering improves discrimination when 5 or more measures are used. Resampling to 102m decreases discriminatory power over all variables tested.

Discriminant analysis is a useful tool for understanding the relationships between several variables and discrimination of ice classes. It is most often used in an exploratory fashion to identify and describe patterns that may exist in the data. The conclusions presented here are the result of the exploratory side of MDA and SDA. Future work will be directed towards the use of more easily computed descriptive variables (first order statistics and image transforms) used in conjunction with texture statistics. Results presented here will be quantified using classical univariate and multivariate hypothesis testing, to determine the stability and universality of these results.

Acknowledgements.

Thanks to T. Carrieres, Atmospheric Environment Service, Ice Branch, for providing ground confirmation, and to J. Piwowar and P. Howarth for review of the manuscript. This research was supported by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science, and an NSERC Operating Grant to E. LeDrew.

Literature Cited

1. Durand, J.M., B.J. Gimonet and J.R. Perbos. SAR Data Filtering for Classification. IEEE Trans. Geos. and Rem. Sens. Vol. 25 (5). pp. 629-637.1987.
2. Haralick R.M., Statistical Image Texture Analysis. Handbook of Pattern Recognition and Image Processing. Ch. 11. Academic Press. 1986
3. Johnston, R.J. Multivariate Statistical Analysis in Geography. Longman, London. 1980.
4. Legendre L. and P. Legendre. Numerical Ecology. Elsevier. Amsterdam. 1983.
5. Ludwig, J.A. and J.F. Reynolds. Statistical Ecology. John Wiley and Sons New York. 1988
6. Manly, B.F.J. Multivariate Statistical Methods. Chapman and Hall, London. 1986.
7. Neter J. and W. Wasserman. Applied Linear Statistical Models. Richard D. Irwin, Inc. Homewood, Illinois. 1974.
8. Shanmugan, K.S., V. Narayanan, V.S. Frost, J.A. Stiles and J.C. Holtzman. Textural Features for Radar Image Analysis. IEEE Trans. Geos. Rem. Sens. Vol. GE, 19, No. 3, pp. 153-156. 1981.
9. Ulaby F.T., R.K. Moore and A.K. Fung. Microwave Remote Sensing: Active and Passive. Vol III. Addison-Wesley Publishing Company. Massachusetts. 1986.
10. Williams B.K. Some Observations on the Use of Discriminant Analysis in Ecology. Ecology. Vol. 64, no.5, pp1283-1291, 1983.

TEXTURE MEASURES FOR SEA-ICE CLASSIFICATION FROM RADAR IMAGES

MOHAMMED E SHOKR

Atmospheric Environment Service
Toronto Canada

(I) ABSTRACT

Automated sea-ice classification from radar images has been identified as one of the prime goals of future sea-ice monitoring programs. The performance of any image classification scheme, supervised or unsupervised, depends on the set of image properties chosen to characterize the existing classes. Therefore, further understanding of those properties that promise strong interpretation capabilities for radar images is highly desirable at present. This paper reports on preliminary results from a recent study to evaluate the utility of selected textural measures, derived from the gray-level co-occurrence matrix, in sea-ice classification using X- and L-band SAR images. The sensitivity of the measures to different computational parameters were also examined.

Key words: sea-ice, texture measures, SAR.

(II) INTRODUCTION

Imaging radars, specially Synthetic Aperture Radar (SAR), have demonstrated a great potential in sea-ice monitoring programs. With the prospect of SAR on future satellites (ERS-1, RADARSAT and EOS), a vast volume of imagery data is expected to be received routinely. To ensure efficient utilization of the data, several research programs have been undertaken to devise automated methods of extracting information from the images especially on sea-ice classification.

An efficient sea-ice classification scheme, the long term objective of the current study, entails defining a suitable set of image properties that separates different ice types uniquely. The two sets commonly used in describing image information are the spectral and textural properties. Spectral properties describe the gray tone as related to electromagnetic radiation received by the imaging sensors, while textural properties describe the spatial distribution of the gray tone within a reasonably small block of pixels. Quantitative definition of texture is rather more difficult, however a few indicators of the gray-tone spatial distribution are usually used to quantify texture. Among those indicators are the degree of coherence (pattern-likeness), homogeneity and contrast. They all contribute to what is commonly perceived as smooth or rough texture.

Gray tone from a single-band radar proved to be not useful for ice classification, and the utility of multispectral radar data remains to be evaluated when such data become more available. Recent studies on SAR image interpretation (Guindon et al., 1982, Holmes et al., 1984, Frost et al., 1984 and Ulaby et al., 1986)

have pointed out the importance of textural information, and suggested combining spectral and textural information for image classification. Several textural measures have been successfully used in previous studies. Measures that are related to solving texture discrimination problems were reviewed (Davis, 1979).

Two main approaches have been adopted in the literature to characterize texture: the first- and second-order texture measures. The first-order measures operate on the gray level of pixels within a specified window to calculate statistical parameters that describe texture. This approach was used by Guindon et al. (1982), to derive and examine three texture measures called homogeneity, contrast and correlation. They concluded that the measures were generally useful in separating rough from smooth ice, and the contrast measure was the principal source of classification improvement. The second-order texture measures operate on a matrix derived from the gray tone values of a window of pixels to calculate statistical parameters from the matrix entries. The parameters can be related to the actual texture content of the window. A widely-used matrix for that purpose is the Gray-Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973). It has been frequently used in SAR image analysis for a variety of applications, but not very much considered for the sea-ice application. An early study (Holmes et al., 1984) examined two GLCM textural measures for sea-ice SAR images: the inertia and entropy. They concluded that first-year ice and multi-year ice could be characterized by different values of inertia texture, while entropy texture could be used to characterize ice floe boundaries and regions of rough-surface. Two more recent studies (Hirose et al. 1989 and Barber et al., 1989) have addressed the application of GLCM texture measures to sea-ice radar images.

The immediate purpose of the present investigation was to examine the utility of selected textural measures (single and combined) for sea-ice classification into four main classes: multi-year (MY), first-year (FY), young ice (YI) and new ice (NI). The dependance of the measures performance on a few computation parameters (Sec. III) was also examined. Results would contribute to a long-term objective of establishing a Comprehensive Radar-Image Properties Library (RIPLIB) that describe the variation of image properties for different ice types under different seasonal and regional conditions. The library can be consulted to select a best set of properties that suits sea-ice SAR interpretation tasks in general, and classification in particular.

(II) DATA SET

Two SAR images were used in this study. The first, Fig. 1, was of Mould Bay in the Canadian North West Territory, obtained in March 1984 using the airborne Inter STAR-1 SAR (X-band, HH polarization). The image was gathered during one of the North-South transects of the aircraft, and was resampled to 35x35 m from 6x6 m pixels. A median filter was applied later to reduce the speckle. The image was made of three ice types: multi-year (MY), first-year rough (FYR), and first-year smooth (FYS). They could all be easily identified manually. Detailed set of field validation data was available.

The second image, of Beaufort Sea/Banks Island Marginal Ice Zone, was obtained in October 1978 from SEASAT SAR (L-band, HH). The image, Fig. 2, was resampled to 50x50 m from 12.5x12.5 m pixels. It was predominantly made up of MY floes (with numerous ridges and rubble) surrounded by a very close pack New Ice (NI) and Young Ice (YI) near shore. New and young ice are remarkably darker than multi-year ice, although YI is slightly lighter than NI and has more texture definition. Training data sets for each ice type in the two images were created based on manual interpretation by sea-ice experts. Mould Bay and Beaufort Sea/Banks Island images will be referred to in this paper as MB and B/B image respectively.

(III) TEXTURE ANALYSIS

The set of textural measures used in this study consists of three statistics derived from the GLCM, namely the inertia, uniformity, and entropy. Definitions of those measures, along with a brief explanation of their relations to the actual textural structure of the image are presented in this section. Main aspects of the GLCM calculations are introduced first.

Let $G(x,y)$ be a gray level function of an image, defined at location (x,y) , and let N be the number of quantized gray levels in G . The GLCM is a square matrix of dimension N . A window of image pixels, Fig. 3, has to be specified in order to compute the matrix entries (in this study a square window of dimensions $W \times W$ was used). The matrix may be assigned to an individual pixel, usually the center of the window, or to the entire block of pixels that comprise the



Fig. 2, SEASAT Beaufort sea image

window (again, in this study a matrix was assigned to each individual pixel). The un-normalized entry P_{ij} of the matrix represents the number of occurrences i,j of two neighbouring pixels within the window, one with gray-level i and the other with gray-level j . The two neighbouring pixels have to be separated by a displacement vector \underline{D} .

$$P_{ij}(\underline{D}) = \# \{ [(x_1, y_1), (x_2, y_2)] \mid G(x_1, y_1) = i, G(x_2, y_2) = j, [(x_1, y_1) - (x_2, y_2) = \underline{D}] \} \quad (1)$$

where $\#$ denotes the number of elements in the set. Three computation parameters have to be selected in order to perform the computation: N , W , and \underline{D} .

Four co-occurrence matrices were computed at each pixel, one for each of the four discrete angles θ associated with the selected value of \underline{D} , i.e. for $\theta = 0, 45, 90$, and 135 (see Fig. 3). The entries were then normalized by dividing by the total number of paired occurrences (varies for each direction) in the window. Normalized entries P_{ij} were calculated as follows (D = magnitude of \underline{D}):

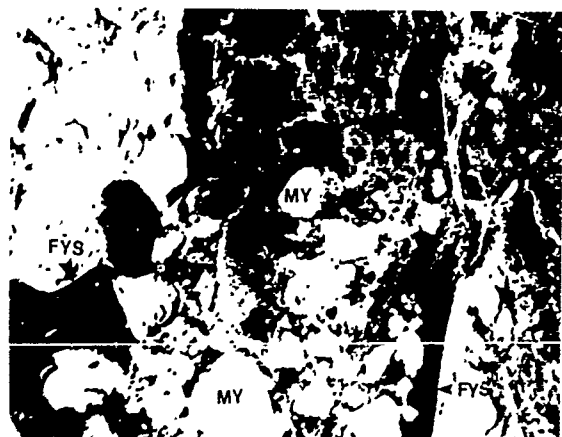


Fig. 1 Mould Bay STAR-1 image

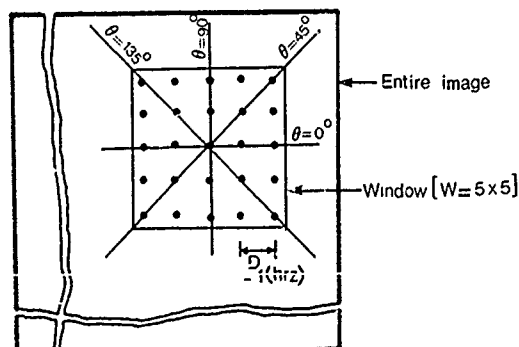


Fig. 3, Geometrical arrangement for GLCM calculation at the center pixel of the shown window.

$$P_{ij}(D) = P_{ij}(D) / [2 W (W-D)] \quad \text{for } \theta = 0, 90 \quad (2)$$

$$P_{ij}(D) = P_{ij}(D) / [2 (W-D)^2] \quad \text{for } \theta = 45, 135 \quad (3)$$

The four matrices were then averaged, and the average matrix was assigned to the pixel under consideration. The advantage of using this matrix is twofold. First, it is rotationally invariant (all directions are involved in the computation) so that identical results are guaranteed if the image is rotated during a transfer process from one medium to another. Secondly, it avoids using an angular-dependent GLCM that may contain misleading information as far as ice classification is concerned. To explain, angular-dependent GLCM may reveal information about texture biased towards a certain direction, but this information should be regarded as disclosure about directionality of ice features (ridges, rubbles, waves, .. etc.) rather than indicators about ice types. Therefore they should not be employed within an ice classification context. The texture measures examined in this study are derived from the GLCM as follows :

$$\text{Inertia} : I = \frac{1}{i,j} \sum_{i=1}^N \sum_{j=1}^N (i-j)^2 \cdot P_{ij} \quad (4)$$

$$\text{Uniformity} : U = \sum_{i=1}^N \sum_{j=1}^N P_{ij}^2 \quad (5)$$

$$\text{Entropy} : E = \sum_{i=1}^N \sum_{j=1}^N P_{ij} \cdot \log(P_{ij}) \quad (6)$$

Inertia and uniformity are sometimes used in the literature under different names: contrast and homogeneity respectively. The definition of inertia is slightly different from the commonly-used version (Holmes, 1984) in that it is normalized by the factor (i,j) , the multiplication of the coordinates of the centroid of the GLCM entries. The coordinates are defined as:

$$\bar{i} = \frac{\sum_{i=1}^N \sum_{j=1}^N i \cdot P_{ij}}{\sum_i \sum_j P_{ij}} \quad (7)$$

$$\bar{j} = \frac{\sum_{i=1}^N \sum_{j=1}^N j \cdot P_{ij}}{\sum_i \sum_j P_{ij}} \quad (8)$$

This normalization procedure was adopted to ensure the invariance of the inertia measure under two possible conditions: linear gray level transformation and re-quantization of the gray level scale. To explain, suppose that same image is acquired from two sensors calibrated differently. The impact on the GLCM will be such that the structure is preserved though displaced in the matrix space. The calculated inertia measure will be affected unless normalized by a factor which reflects that displacement. A similar argument applies when GLCM is calculated for the same image using different N values. In this case, GLCMs will have different sizes and the spatial distribution of entries will be compressed or enlarged accordingly. Again, the inertia measure will be affected unless normalized by a factor that involves the centroid of the modified distribution.

The actual textural structure of an image subscene can be related to the structure of the co-occurrence matrix as follows. For a homogeneous subscene, transitions between high and low gray levels are not frequent. Therefore, most of the pairs of adjacent pixels will have similar gray levels, and that leads to large values of diagonal and near-diagonal entries of the matrix. On the other hand, if a subscene is highly textured, transitions between high and low gray level occur more frequently, and that builds towards large values of the off-diagonal entries.

Next to this relation is the relation between the structure of the co-occurrence matrix and the values of the texture measures defined above. Inertia is highly influenced by the off-diagonal values of the matrix. Therefore, it should be considered as a measure of contrast between the gray levels in the computation window. High inertia values mean higher contrast texture. Uniformity, on the other hand, is mainly influenced by high-value entries of the matrix. Bearing in mind that such high values result when the gray level distribution over the window has either a constant or a periodic form, then uniformity as the name indicates, can be considered as a measure of the homogeneity of gray level distribution. A high uniformity value means more homogeneous region and vice versa. Entropy, as pertaining to the random process theory, is a statistical measure which is used to quantify the "disorder" in a function. To illustrate, consider two extreme cases of the gray level function: a uniform, and a fully random (white noise) function. In the first case, the histogram is a Delta function and therefore all but one of the P_{ij} in the entropy equation are zero, and the entropy achieves a minimum value of zero. In the second case, the histogram is a constant function, i.e. all P_{ij} are equal, and equal to $1/m$ where m is the total number of entries in the signal. In this case the entropy reaches its maximum value of $\log m$ (Whittle, 1970). In conclusion, the entropy takes a higher value as the histogram of the gray level distribution becomes closer to uniform (more disordered).

Computation of the three texture measures at each pixel resulted in three "texture images". The gray level at each pixel in a texture image is proportional to the mean value of the relevant measure. Texture images can be generated in a reasonable computation time only if the gray level range is reduced from the commonly-available 255 intervals (8-bit). For that reason, the gray level was compressed to 4, 16 and 64 intervals. The relevant computation time for any texture image was 8, 45, and 470 minutes respectively on the Concurrent 3244 computer of the Aerospace Meteorology Division of Atmospheric Environment Service (AES). A special application program was developed under PCI EASI/PACE image analysis software to perform the computation. The gray level compression was achieved by examining the histogram of the raw image and re-quantizing the gray level band centered on the mean value and bounded by a width of ± 3 standard deviations.

(IV) RESULTS

All values of texture measures presented in this section are in the actual "texture-measure" scale rather than "texture-image" scale. The difference between the two scales is that the latter (usually 8-bit) is derived from the former (real numbers typically between 0.0 to 1.0) using a linear mapping function in order to allow the display of a texture image. Obviously, the mapping function depends on the

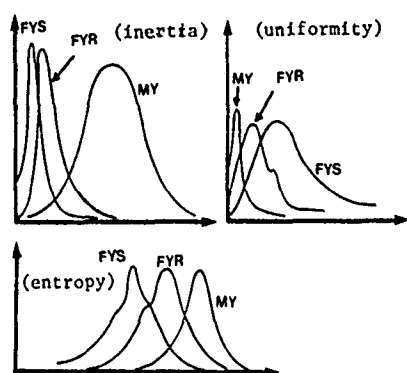


Fig. 4 Histograms of inertia, uniformity and entropy from Mould Bay image.

Table I GLCM parameter values used in the study

| Examined parameter | values of the other two parameters | | |
|--------------------|------------------------------------|---|---|
| | N | W | D |
| N=4,16,64 | — | 5 | 1 |
| W=5, 7, 9 | 16 | — | 1 |
| D=1, 2, 3 | 16 | 7 | — |

image content, and therefore texture values obtained in the texture-image scale from a certain image will lose their meaning when compared to corresponding values from another image (comparison between texture from different ice types in the same image should still be possible). Texture-measure scale has to be used if the purpose is to assimilate data from several test images to define suitable discriminating measures applicable to any SAR image.

Histograms of different texture measures for each ice type were evaluated for their use in characterizing ice types. No characteristic distributions were found although both the inertia and entropy of MY ice tended to fit a Gaussian function. Distributions of all texture measures for YI and NI had a large peak followed by a long, thin tail at one side. The uniformity measure was characterized by blunt tails. Selected distributions are shown in Fig. 4 for MB image. No attempt was made to construct and test hypotheses about histogram of a population (an ice type) as related to the histogram of a sample (training data of an ice type).

Sensitivities of the texture measures to the three GLCM control parameters: N, W and D, were examined.

Table I shows the parameter values used for that purpose. The mean values of the three measures with error bars representing \pm one standard deviation from means, are plotted in Figs. 5 for the three ice types in the MB image. It is important to interpret the error bars in relation to the sample distribution functions described above.

As can be seen from the figures, the sensitivity of texture measures to N is tangible only between N=4 and 16, after which no significant change is noticed. This means that there is no need to perform texture calculation for $N > 16$, which is desirable from the standpoint of computation time. Results from the B/B image suggests that texture measures calculated using N=4 may be used to furnish a preliminary discrimination between MY ice and other ice types. This may be particularly useful for a near-real-time ice classification scheme. With respect to the window size and pixel displacement, texture measures are almost insensitive. A minor exception here is the inertia, which demonstrates a slightly less discrimination capability between ice types as the window size increases. This observation is evident in the results of both images. Results from Figs. 5 also suggest the selection of N=16 (4-bit image), W=5X5 and D=1 as the best set that preserves the texture information while ensuring a reasonable computation time. These values were used to construct table II, where gray level and texture-measures statistics (mean, median, and standard deviation) are shown for different ice types in the MB and B/B images. The table constitutes a small portion of the proposed RIPLIB, defined in Sec. II.

A few interesting observations can be drawn from the table. First, MY ice in both images had almost identical textural statistics, notwithstanding the differences in radar frequencies (X-band for MB and L-band for B/B), the region and the season under which the two images were acquired. This is a promising result because it confirms the possibility of identifying a region-independent classifier parameter. It is worth mentioning that texture statistics, in terms of texture-image scale, of MY ice from both images, were not identical as such. The second observation is pertinent to the significant difference between the gray level statistics from the two images. MY ice is nearly twice as bright in B/B image as in MB image. The difference may be attributed to the question of SAR radiometric calibration. As a result, gray level should not be used to characterize ice types (in an absolute sense) unless a proper normalization procedure is devised. Here again, the invariant character of the GLCM (and the derived textural measures) under monotonic gray level transformations proved to be a highly desirable property. Thirdly, the table indicates the difficulty in discriminating FYR, FYS, and YI. The YI has a texture closer to the FYR ice. Additional image properties are required to clarify the fuzzy region occupied by those three ice types.

Table II Statistics of gray tone and texture measures for different ice types

| ice type | DATA | Gray tone | | | Inertia | | | Uniformity | | | Entropy | | |
|----------|------|-----------|-----|-------|---------|--------|--------|------------|--------|--------|---------|--------|--------|
| | | MEAN | MDN | SD | MEAN | MDN | SD | MEAN | MDN | SD | MEAN | MDN | SD |
| MY | MB | 57.38 | 56 | 11.56 | 0.1923 | 0.1725 | 0.0863 | 0.0307 | 0.0265 | 0.0131 | 0.8101 | 0.8083 | 0.0603 |
| MY | B/B | 120.30 | 120 | 30.63 | 0.1908 | 0.1745 | 0.0828 | 0.0290 | 0.0264 | 0.0121 | 0.8261 | 0.8259 | 0.0594 |
| FYR | MB | 28.53 | 27 | 7.84 | 0.0562 | 0.0531 | 0.0271 | 0.0911 | 0.0806 | 0.0437 | 0.6431 | 0.6448 | 0.0903 |
| FYS | MB | 22.33 | 22 | 5.73 | 0.0280 | 0.0265 | 0.0172 | 0.2061 | 0.1805 | 0.1131 | 0.4866 | 0.4991 | 0.1331 |
| YI | B/B | 44.83 | 44 | 15.75 | 0.0445 | 0.0388 | 0.0281 | 0.1308 | 0.1221 | 0.0618 | 0.5880 | 0.5830 | 0.0945 |
| NI | B/B | 11.66 | 8 | 16.07 | 0.0205 | 0.0097 | 0.0363 | 0.5038 | 0.5091 | 0.3018 | 0.3051 | 0.2694 | 0.2100 |

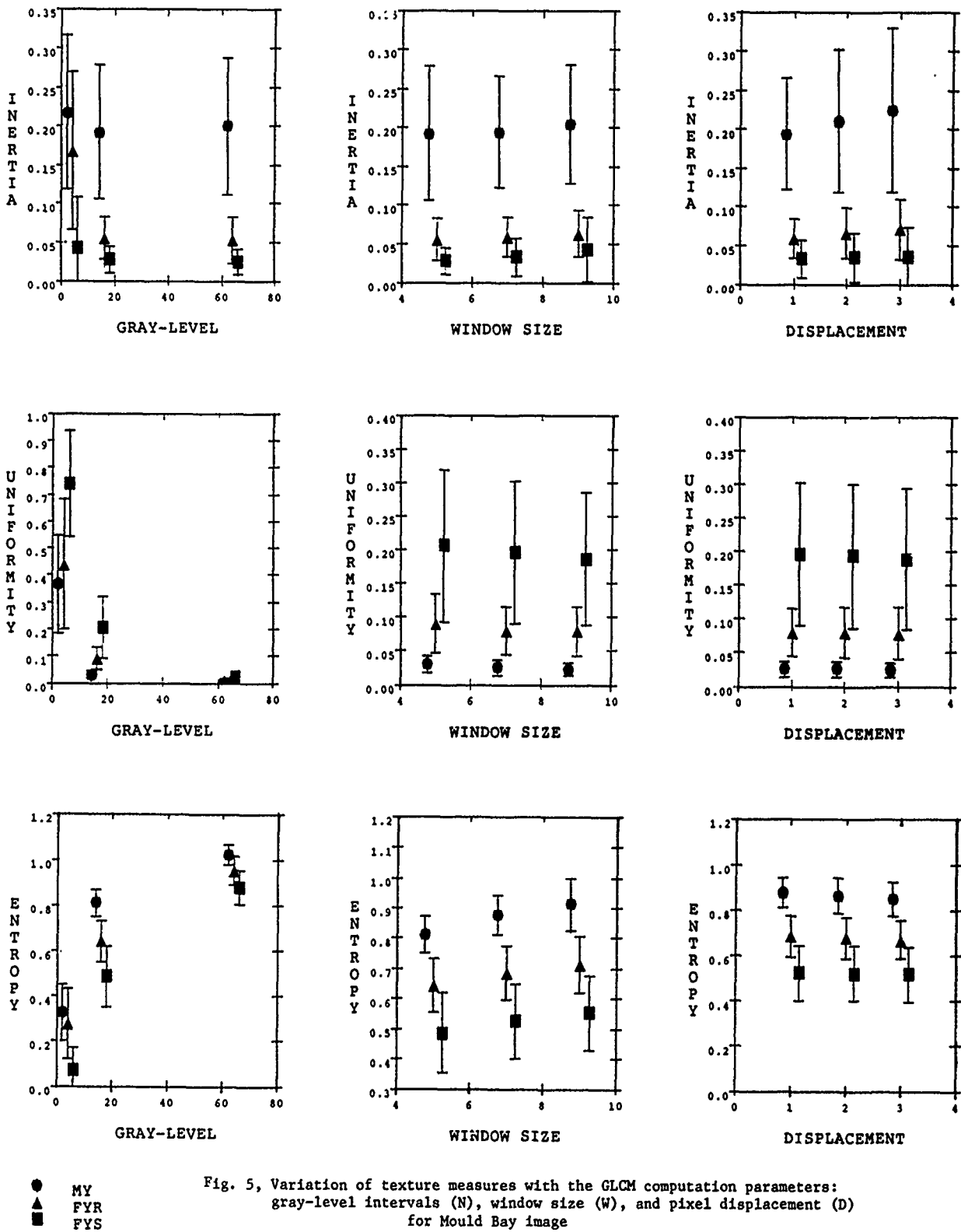


Fig. 5, Variation of texture measures with the GLCM computation parameters: gray-level intervals (N), window size (W), and pixel displacement (D) for Mould Bay image (error-bar lines represent \pm one standard deviation)

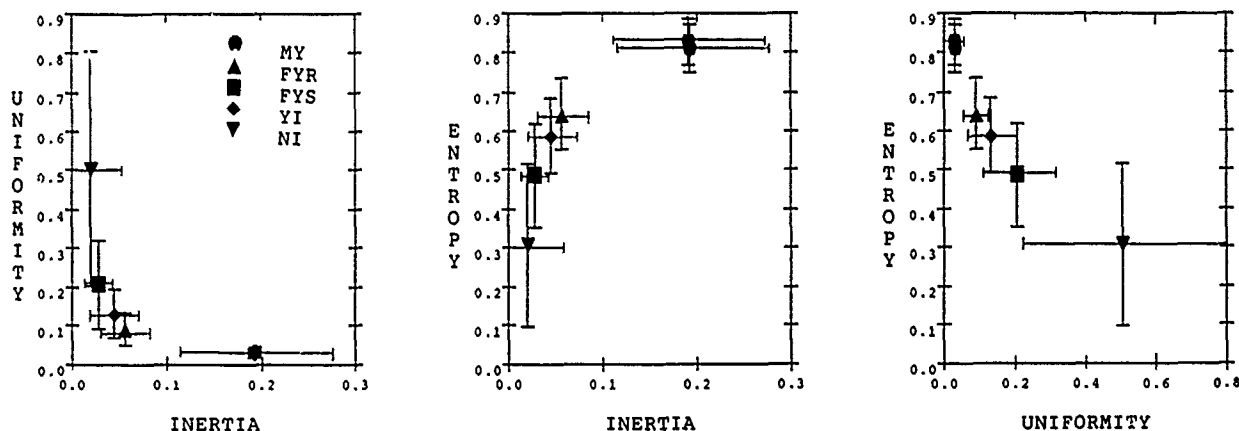


Fig. 6, Variation of texture measures for different ice types in 2-D feature space

Combinations of textural measures are plotted in two-dimensional feature space as shown in Fig. 6. The plots are for $N=16$, $W=5$ and $D=1$. Qualitative examination of the plots allows preliminary evaluation of the performance of a given combination in a classification routine. They indicate that MY ice can be discriminated in any space, while NI is relatively identifiable in the uniformity-entropy space. Because of the long-tailed histogram of texture measures of the NI, a measure variance can be significantly reduced if a small portion (e.g. 5%) of the tail is eliminated, and therefore the NI separability can be improved. The figures also show that the uniformity measure is more sensitive to smooth texture, (small variance for MY and large variances for YI and NI). No similar trend can be noticed for the inertia and entropy measures. Inertia and entropy offer same expression for ice types, while uniformity offers an inverse expression.

(V) CONCLUSIONS

This study was part of a scientific endeavor undertaken by a Canadian group, CRAGTAG, to develop an automated algorithm for sea ice classification from SAR images. The adopted strategy entailed defining radar-image properties that could separate ice types uniquely. To that end, the present investigation intended to evaluate the utility of three texture measures, derived from gray-level co-occurrence matrix, as sea-ice classifiers. Within the limitations of the analyzed data, the following four main conclusions were attained:

1. The examined texture measures are not sensitive to the three GLCM computation parameters (N , W , and D). A 4-bit gray level scale is sufficient to preserve all texture information required for the analysis.
2. Multi-year ice can be identified using any measure although minimum variance is offered by the uniformity measure. New ice can also be best identified by the uniformity.
3. None of the examined measures is useful in separating first-year smooth and rough, and new ice.
4. GLCM texture measures are of more universal character than gray tone, in that an ice type can be associated with a unique range of a texture measure. This is due to the invariant character of the GLCM under monotonic GL transformation.

ACKNOWLEDGEMENT

The author wishes to acknowledge the contribution of AES Ice Branch in interpreting the SAR images used in the study.

References:

1. Barber, D.G. and LeDrew, E., "Multivariate analysis of texture algorithms for SAR sea ice classification", Proc. Int. Geosci. Remote Sensing Symposium, Vancouver, Canada, 1989 (to appear).
2. Davis, L., "Image texture analysis techniques - A survey", Digital Image Processing, D. Reidel Publishing Co., pp. 189-201, 1981.
3. Frost, V.S., Shanmugan K.S. and Holtzman J.C., "The influence of sensor and flight parameters on texture in Radar Images", IEEE Trans. Geosci. Remote Sensing, Vol. GE-22, No.5, pp 440-448, 1984.
4. Guindon B., Hawkins, R.K. and Goodenough D.G., "Spectral-spatial analysis of microwave sea ice data", Proc. Int. Geosci. and Remote Sensing Symposium, 1982.
5. Haralick, R., Shanmugan K. and Dinstein I., "Textural features for image classification", IEEE Trans. Syst. Man, and Cybernetics, Vol. SMC-3, pp. 610-621, 1973.
6. Hirose, T., Paterson, L. and McNutt, L., "On the use of texture to augment sea ice classification", Proc. Int. Geosci. Remote Sensing Symposium, Vancouver, Canada 1989, (to appear).
7. Holmes, Q.A., Nuesch, D.R. and Shuchman R.A., "Textural Analysis and real-time classification of sea-ice types using digital SAR data", IEEE Trans. Geosci. Remote Sensing, Vol. GE-22, No.2, pp.113-120, 1984.
8. Ulaby, F.T., Kouyate, F., Brisco B. and Williams T.H., "Textural information in SAR images", IEEE Trans. Geosci. Remote Sensing, Vol. GE-24, No.2, pp.235-245, 1986.
9. Whittle, p., "Probability", Penguin Books Ltd, p. 218, 1970.

COMPARISON OF SEA ICE PARAMETERS RETRIEVED FROM PASSIVE MICROWAVE (SSM/I), LANDSAT MSS AND AVHRR IMAGERY

Axel J. Schweiger, Cooperative Institute for Research in Environmental Sciences
University of Colorado, Campus Box 449, Boulder, Co 80309
Email: kryos::schweig (SPAN); schweiger@colorado (BITNET)

Konrad Steffen, Swiss Federal Institute of Technology, ETH
Dept. of Geography, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
Email: K.Steffen/Omnet (Telemail); geograph@czheth5a (BITNET)

ABSTRACT

Ice concentrations derived from passive microwave data from the Special Sensing Microwave Imager SSM/I, Landsat-Multispectral Scanner (MSS) and Advanced Very High Resolution Radiometer (AVHRR) are intercompared for the validation of SSM/I derived sea ice products. Case studies with data from the Beaufort, Bering, East Greenland and Weddell Sea are conducted to evaluate the performance of the NASA - Team SSM/I algorithm under various sea ice conditions. The results show that SSM/I ice concentrations are accurate with a mean error of 15 % using globally derived tie points.

INTRODUCTION

Polar oceans play an important role in the global climate system. Recent modelling results indicate that the warming effect of "greenhouse" gases in the atmosphere will be amplified in the polar regions where the presence of sea ice provides a positive feedback (Hansen, 1988). Further studies investigating the role of the polar regions in global climate change require a better understanding of heat and mass transfer processes in the polar oceans. These processes are largely controlled by the geographical distribution of sea ice as well as types. Passive microwave remote sensing techniques with the ability to determine first and multi-year ice concentrations have been used to monitor sea ice in the arctic regions since 1973 (Cavalieri et al., 1984; Comiso, 1986; Gloersen and Cavalieri, 1986; Steffen and Maslanik, 1988). With the launch of the Special Scanning Microwave Imager (SSM/I) on board a Defense Meteorological Satellite Program (DMSP) passive microwave data have become available that allow the compilation of a long term data base for the sea ice research community. The research presented is part of an ongoing effort to validate and improve sea ice products derived from SSM/I brightness temperatures through comparison with higher resolution LANDSAT-MSS and AVHRR imagery. It addresses the general question to what accuracy sea ice concentrations can be determined in different geographical regions with varying sea ice regimes.

METHODOLOGY

The accuracy assessment of sea ice products based on microwave radiometry is significantly complicated by the large fields of view of the SSM/I passive microwave sensor (69x43 km for the 19 GHz and 37x28 km for the 37 GHz channel). In effect the sensor integrates over areas substantially larger than most ice features. Therefore, higher resolution imagery, allowing the classification of ice features on a pixel level, needs to be compiled for areas covering several SSM/I footprints in order to allow a meaningful comparison. Since a comparison at the hemispheric scale typical for the SSM/I sea ice products is impossible, case studies have to be designed to comprise a representative sample reflecting the geographical and seasonal variability of sea ice conditions in the north and south polar

regions. The presented case studies from the Beaufort, Bering, Weddell and East Greenland Sea reflect ice conditions during freeze up and spring with low and high ice concentrations.

Ice concentrations are determined from the higher resolution imagery, averaged over 50 by 50 km grid cells corresponding to the approximate size of the SSM/I 19 GHz channel footprint and compared to sea ice parameters calculated from the SSM/I brightness temperatures. The analysis of errors then reveals possible problems with the SSM/I algorithm with respect to a particular ice regime and the selection of algorithm parameters such as the assumed brightness temperatures for first- and multi-year ice and open water (tie points).

DATA PROCESSING

Landsat MSS

Cloud-free Landsat imagery was acquired for a range of ice conditions and geographic locations to allow for a validation of the SSM/I sea ice algorithm for a variety of ice concentrations and ice types. Landsat imagery was acquired in two formats: Landsat MSS digital imagery and Landsat MSS photographic transparencies.

Digital imagery

Geometrically and radiometrically corrected Landsat MSS imagery at 80 m resolution (Field of View) is available from EOSAT corporation. This imagery can easily be geolocated and regridded to match the SSM/I grid format (polar stereographic projection). Due to the high reflectance levels of ice, MSS channels 1-3 are frequently saturated and can only be utilized in the ice parameter classification under low sun angle illumination conditions. However, reflectance differences between thin ice types and open water are greatest in the near infrared spectral range, so that MSS channel 4 (0.8-1.1 μ) is most suitable for the determination of ice concentrations.

Photographic transparencies

Due to data communication problems with the satellite downlink (TDRS), no digital imagery is available for some areas in the Bering, Chukchi and Beaufort Sea, but photographic transparencies could be ordered through the Alaskan Quicklook system operated by the University of Alaska's Geophysical Institute (Miller et al., 1981). This imagery is received in real time and recorded on photographic transparencies. Data are geometrically corrected for earth rotation but geolocation information is unreliable due to the lack of accurate ephemeris and real time satellite attitude data. Landsat MSS channel 4 data were therefore acquired in swaths so that at least one frame within each swath contains identifiable landmarks. Through comparison with maps and digital coastline data (CIA World Data Bank 2), a geolocation correction can be calculated for the entire path resulting in a residual geolocation error of 1.5 km with respect to the map data. For the subsequent determination

of ice concentration and sea ice types the transparencies were digitized using a high resolution scanner, geolocated using the above outlined procedure and projected to a polar stereographic projection (true at 70° latitude) used for the SSM/I gridded sea ice products.

Sensor Noise Removal

Landsat MSS image reflectance values are very low due to the low illumination conditions during spring and fall in the Arctic. Despite the high albedo of sea ice surfaces, images frequently display maximum digital numbers (DN) not exceeding 40. Signal to noise ratios therefore are enhanced and significant 'stitching' interference patterns become noticeable on contrast stretched imagery (Miller, 1986). A fast fourier transform (FFT) filter was used to remove sensor noise which made the imagery more suitable for classification and extraction of sea ice concentrations.

Effects of varying solar zenith angles on Landsat imagery

During the process of classification of the Landsat MSS imagery, low frequency brightness variations, a darkening from one side of the image to the other, were discovered on many of the images. These brightness variations seem to be unrelated to ice type variations or cloud interference but rather consistently occurred along an axis in the direction of the sun. The brightness variation was strong enough to influence the result of the brightness threshold based sea ice type classification scheme. Fig. 1 shows the relationship between the variation of solar elevation above horizon across the image, and the mean DN values (only values above DN = 100 were sampled). Despite the small solar elevation changes of only 2.3 degrees, a substantial change in mean DN values can be observed. Snow and ice surfaces display strong forward scattering characteristics at low sun angles (Steffen, 1987; Taylor and Stowe, 1984), thereby reducing the radiance in the direction of a near nadir satellite. In light of previous research, the magnitude of the solar zenith angle dependence is surprising. Nonlinearities during the film recording and digitization processes prevent an absolute calibration of DN values to radiances for the redigitized products. To reduce this effect for the purpose of ice classification, images displaying a large brightness variation due to the solar zenith effect are corrected using a linear fit of observed mean brightness values to the solar elevation.

Ice classification

For the determination of ice concentration MSS channel 4 is most useful, since it displays the greatest reflectance differences between thin ice types and open water. The capability to distinguish thin ice types seems crucial for the validation of the SSM/I sea ice parameters since sea ice concentrations derived from passive microwave data are too low in the presence of thin ice (Steffen and Maslanik, 1988). A combination of channel 1 and 4 displayed some utility for the classification of clouds and cloud shadows. For the ice type classification, two methods were developed.

Threshold techniques

Using training areas, the brightness value ranges for different ice types and open water are determined. By selecting appropriate brightness thresholds, an image can be classified into 5 different ice types. For the determination of ice concentration the class spectrum comprised open water/dark nilas, light nilas, grey ice, grey-white ice and white ice, corresponding to a categorization of ice thickness and stages of development commonly used in sea ice research (Steffen, 1986). During spring when only open water and white ice are present, a 'subresolution' class was introduced. Pixels at intensity levels between open water and white ice were interpreted as containing ice floes of subresolution sizes and put into this class. Following classification, ice concentrations are calculated for 50 by 50 km grid boxes corresponding to the grid of the SSM/I sea ice concentration products.

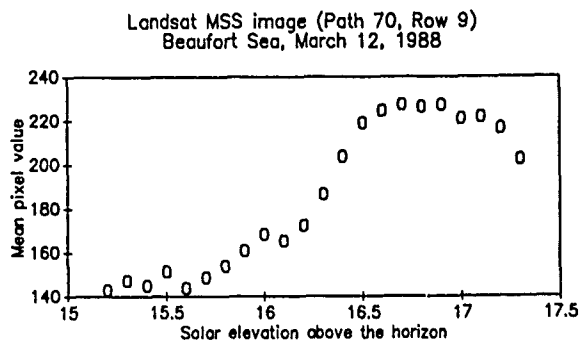


Fig.1 Mean pixel values for a Landsat MSS image in the Beaufort sea vs. Solar elevation above the horizon in Deg.

Tie point algorithm

During periods where no new ice formation occurs, the spectrum of classes is reduced to open water and white ice. If open water and white ice are the only two classes that are present, the assumption can be made that all brightness values in between those classes must represent ice concentrations at sub-resolution. The following algorithm (Comiso and Zwally 1982) was devised to more realistically account for the presence of ice floes smaller than the MSS resolution.

$$Ic = (Dx - Dl / Dh - Dl) * 100$$

where:

- Ic = Ice concentration
- Dx = Brightness value representing ice concentration
- Dl = Brightness value for open water
- Dh = Brightness value for white ice

Tie points were found using training areas for open water and large white ice floes, where Dh represents the mean brightness for that floe minus one standard deviation. The tie point algorithm produces better results due to its capability of accounting for sub-resolution size ice floes. This was also found by Comiso and Zwally (1982) in comparison with electrical scanning microwave radiometer (ESMR) derived sea ice concentrations.

AVHRR

AVHRR local area coverage (LAC) and direct read out images (HRPT) were obtained and registered to a polar stereographic projection for comparison with SSM/I and Landsat data. The tie point ice concentration algorithm developed for Landsat imagery was applied to the AVHRR data. So far no direct comparison of ice concentrations from AVHRR and Landsat data in the visible channels has been carried out. However, it is one of the goals of the ongoing validation effort. Ice concentrations derived from thermal infrared AVHRR data showed substantially lower values compared to the Landsat derived products, and therefore, will not be used in future comparisons.

SSM/I

Passive microwave data used in this study were acquired by the DMSP special sensor microwave imager (SSM/I). The instrument operates at four frequencies: 19.3, 22.2, 37.0 and 85.5 GHz. Vertical and horizontal polarizations are provided for each frequency, except the 22.2 GHz channel. SSM/I orbital data supplied by Nasa Ocean Data System (NODS) and National Snow and Ice Data Center (NSIDC) were mapped to grid cells with a dimension of 25x25 km and projected to a polar stereographic projection (true at 70° latitude). For the calculations of ice concentrations, brightness temperatures were averaged over 50x50 km grid cells.

Ice Classification

Mapping of sea ice concentration using passive microwave data is possible owing to the large difference in polarization between water and sea ice. In addition, the drainage of brine pockets during summer causes alterations in the scattering properties that allow a differentiation between ice types (Svendsen et al., 1983; Comiso, 1983; 1986; Cavalieri et al., 1984; Gloersen and Cavalieri, 1986). Using the algorithm devised by Gloersen and Cavalieri (1986), sea ice concentrations and multi-year ice fractions are calculated from the 19 GHz polarizations and 37 - 19 GHz brightness temperature differences of the SSM/I data. The algorithm interprets measured brightness temperatures as the respective contributions of open ocean, first-year and multi-year ice and assumes fixed brightness temperatures values for open water (T_{BW}), 100 % first-year ice (T_{BF}) and 100 % multi-year ice (T_{BM}). These fixed temperatures, known as tie points are critical for the accurate performance of the ice concentration retrieval algorithm and are commonly determined from statistics for the entire polar ocean (Cavalieri and Gloersen 1984). An analysis of the sensitivity of the retrieval accuracy to the selection of tie points is discussed by Steffen and Schweiger (These Proceedings).

RESULTS

Beaufort Sea

Four case studies were conducted for the Beaufort Sea region covering an ice concentration range from 0 to 100 % during fall and spring. Case study 1 represents the situation of low ice concentration (mean = 13 %) along the ice edge during the fall with air temperatures above freezing. Case study 2 represents the freeze-up situation showing Landsat ice concentration between 91 and 97 % with 5.9 % open water and dark nilas, 5.1 % grey ice, 38.3 % grey-white ice, and 50.3 % white ice. Case studies 3 and 4 represent the very close pack ice situation in spring with ice concentrations above 95 % largely composed of first- and multi-year ice types. Mean Landsat - SSM/I ice concentration differences for the Beaufort sea area case studies range between 7.5 % and 13.9 % with extreme deviations of up to 19.7 %.

Bering Sea

In the Bering Sea (Case study 5) the Landsat ice type classification shows total ice concentrations between 88 and 100 %. The ice regime is dominated by young ice with over 70 % nilas, grey and grey white ice. In the cloud free regions, more than 70 % of the ice cover was young ice (grey ice and grey-white ice) and nilas. Further comparison with high resolution aerial photographs showed that linear features of 120 m in dimension such as ice floes or fractures could still be identified in the Landsat image. Ice concentration differences between Landsat and SSM/I are large in this areas showing a mean error of 15 % percent and a maximal difference of 24 % can be observed. This large discrepancy is most likely due to the inability of the NASA algorithm to detect thin ice types.

Weddell Sea (Case study 6)

Ice regime is close pack ice with concentrations between 8/10 and 9/10, composed of floes mostly in contact. Floes size varied from small (<100 m) to giant (>10 km), with the largest floes being approximately 40 km in diameter. No significant open water areas are present except in leads. The objective of this case study is the validation of high first-year ice concentration during premelt. Landsat and SSM/I derived ice concentrations are in good agreement with a mean difference of 3.7 percent. Detailed results for the Landsat-SSM/I comparison by cases studies are presented in table 1 with separate calculations for 25 and 50 km grids.

| Case | Mean | Std.Dev | Min | Max | N | Grid |
|------|-------|---------|-----|------|----|------|
| 1 | 9.7 | 3.4 | 5.7 | 13.9 | 9 | 50 |
| 1 | 9.15 | 5.8 | .4 | 22.3 | 51 | 25 |
| 2 | 13.9 | 3.9 | 9.9 | 19.7 | 6 | 50 |
| 2 | 15.7 | 6.1 | 1.1 | 24.9 | 36 | 25 |
| 3 | 7.5 | 1.4 | 5.4 | 9.8 | 7 | 50 |
| 3 | 7.8 | 2.1 | 2.1 | 14 | 47 | 25 |
| 4 | 11.1 | 1.2 | 9.5 | 13.1 | 12 | 50 |
| 4 | 11.1 | 2.0 | 6.7 | 15.0 | 64 | 25 |
| 5 | 15.8 | 4.9 | 5.4 | 24.1 | 14 | 50 |
| 5 | 15.13 | 4.8 | 1.9 | 25.8 | 87 | 25 |
| 6 | 3.9 | 1.3 | 2.8 | 6.2 | 7 | 5 |
| 6 | 3.7 | 2.3 | .3 | 11.1 | 34 | 25 |

Table 1. Statistics of Landsat-SSM/I comparison for 2 different grid size. Differences are expressed in absolute mean errors. Case 1,2,3,4 Beaufort Sea Area, Case 5 : Bering Sea, Case 6: Weddell Sea. N is the number of observations per case study.

AVHRR East Greenland Sea

SSM/I derived ice concentrations based on NASA tie point statistics are compared with average AVHRR derived ice concentration for 50 by 50 km size pixels for an area in the East Greenland Sea (Fig. 2). The ice regimes is governed by a mixture of first and multi year ice flows with ice concentrations ranging from 0 to 70 %. The SSM/I ice concentration algorithm overestimates ice concentration on average by 7.6 %. A comparison based on the 25 km shows a mean error of 9.6 % with extreme values of errors are substantially larger than for the 50 km grid (42 % and 24 % respectively). In light of the fact, that this particular case study reflects the more complicated summer situation with potential melt-freeze on top of the ice, these results must be viewed to be good.

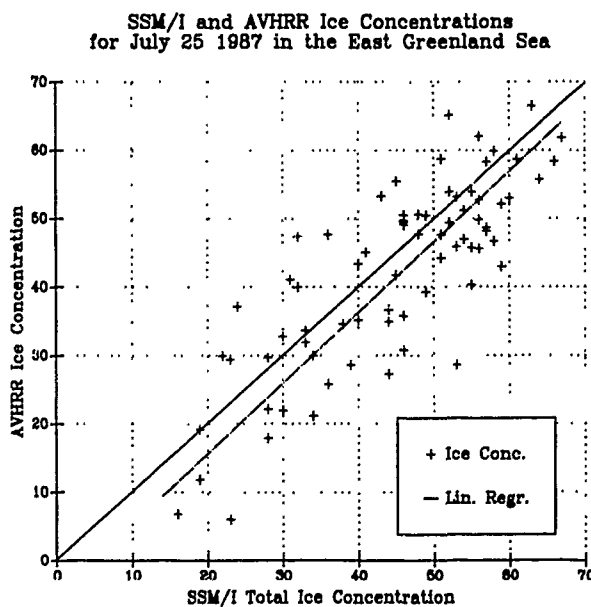


Fig.2 Comparison of AVHRR and SSM/I derived ice concentrations for an area in the East Greenland Sea (July 25, 1987). Solid line represents the line of perfect agreement. Dashed line represents a linear regression of AVHRR on SSM/I ice concentrations.

Weather effects

Effects due to wind or intense cloud systems can substantially alter the brightness temperatures over open ocean and lead to spurious ice concentrations of up to 18 % in ice free areas of the East Greenland Sea AVHRR image. The weather filter derived by Cavalieri and Gloersen (1984) proved ineffective when applied to SSM/I data. Fig. 3 shows a scatterplot of gradient vs polarization ratios for manually classified ice covered areas (any concentration) and areas displaying intensive cloud cover due to a storm system moving through the area. The clear distinction between the cloud and ice clusters allows the filtering of weather contaminated pixels through the application of a gradient threshold at 0.05. This threshold is considerably lower than the one suggested for use with SMMR data and reflects the greater proximity of the 19 GHz channels of the SSM/I to the 22 GHz water absorption band in contrast to the 18 GHz channel of the SMMR. Investigations by other researches support this finding (Cavalieri pers. Comm)

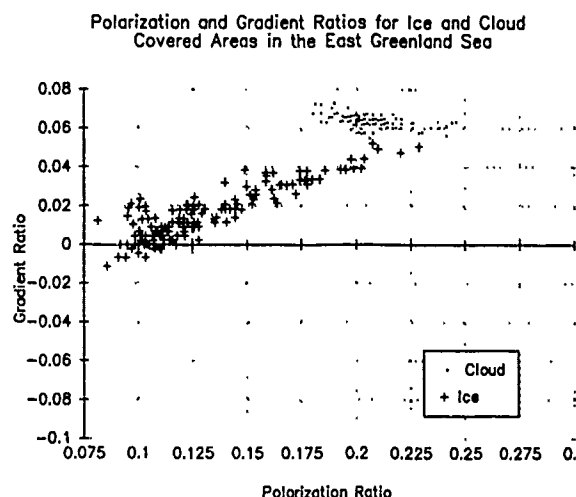


Fig.3 Scatterplot of gradient and polarization ratios for ice covered and ice free but cloud covered areas. Classes are based on a manual interpretation of the AVHRR East Greenland Sea image

DISCUSSION

Mean retrieval accuracies in the 15 % percent range seem reasonable with respect to the low resolution and quality this data set for studies concerned with large scale sea ice distribution eg. hemispheric ice extent. Yet, for more regional studies scale investigation, or for studies requiring the detection of small changes in ice concentrations such as an response to greenhouse warming, a higher accuracy is desirable.

A analysis of the error sources points to the fact that the globally derived tie points are responsible for a substantial amount of the error in some of the case studies. They ignore variations of surface roughness, ice salinity, wave action and ocean spray as well as atmospheric water content. A more detailed analysis of this problem, as well as a method for deriving locally adjusted tie points is presented by Steffen and Schweiger (These Proceedings)

In addition, there is always a time difference of data acquisition between Landsat and SSM/I satellites which accounts for some differences, mainly in areas of low ice concentrations. In the Beaufort Sea, for example, the time lag between the two satellites was 6 hours.

The greater accuracy of the 50 km grid ice concentrations is related to the different fields of view of the 19 and 37 GHz channels, 69x43 km and 37x28 km respectively. For the 19 GHz channel gridded to a 25 km grid, only about 50 % of the radiant energy originate from the surface covered by that particular grid cell. The remaining 50 % are emitted from adjacent surface areas

outside the cell. In regions, of large brightness temperature contrast, such as along the ice edge, this can lead to spurious ice concentrations and therefore to large discrepancies in the comparison with high resolution sea ice products. This hypothesis is supported by the AVHRR-SSM/I comparison which shows the largest discrepancies in ice concentrations along the ice edge. It is therefore recommended to calculate 50 km brightness temperature averages for the 19 and 37 GHz channels prior to the calculation of ice concentrations.

REFERENCES

- Cavalieri, D. J., P. Gloersen, and W. J. Campbell, "Determination of sea ice parameters with Nimbus 7 SMMR", *J. Geophys. Res.*, Vol.89, pp. 5355-5369, 1984.
- Comiso J.C., and H.J. Zwally, "Antarctic sea ice concentrations inferred from Nimbus 5 ESMR and Landsat" Imagery, *J. Geophys. Res.*, Vol.87, No.C8, pp. 5836-5844, 1982.
- Comiso, J. C., "Sea ice effective microwave emissivities from satellite passive microwave and infrared observations", *J. Geophys. Res.*, Vol.88, No.C12, pp. 7686-7704, 1983.
- Comiso, J. C., "Characteristics of arctic winter sea ice from satellite multispectral microwave observations", *J. Geophys. Res.*, Vol.91, No.C1, pp. 975-994, 1986.
- Gloersen, P., and D. J. Cavalieri, "Reduction of weather effects in the calculation of sea ice concentration from microwave radiances", *J. Geophys. Res.*, Vol.91, No.C3, pp. 3913-3919, 1986.
- Hansen, J., I. Fung, A. Lacis, D. Rind, S. Lebedeff, R. Ruedy, G. Russell, and R. Stone, "Global climate changes as forecasted by Goddard Institute for Space Studies three-dimensional model", *J. Geophys. Res.* Vol. 93, No.D8, pp. 9341-9364, 1988.
- Miller, J.M., Campbell, N., Mackinnon, R., "An Experimental Landsat Quick-Look System for Alaska", *Proc. Seventh International Symposium on Machine Processing of Remotely Sensed Data*, pp. 639-646, 1981.
- Miller, J.M., Berger, G.B., "Enhancing Landsat Data Acquired under very low illumination", *Photogrammetric Engineering and Remote Sensing*, Vol. 52, No.6, June 1986, pp. 801-807, 1986.
- Steffen, K., "Ice conditions of an Arctic polynya: North Water in winter", *J. Glaciol.*, Vol.32, No.112, pp. 383-390, 1986.
- Steffen, K., "Bidirectional reflectance of snow at 500-600 nm", In: *Large Scale Effects of Seasonal Snow Cover*, *IAHS Publ. No.166*, pp. 415-425, 1987.
- Steffen, K., and J.A. Maslanik, "Comparison of Nimbus 7 scanning multichannel radiometer radiance and derived sea ice concentrations with Landsat imagery for the North Water area of Baffin Bay", *J. Geophys. Res.*, Vol.93, No.C9, pp. 10769-10781, 1988.
- Svendsen, E., K. Kloster, B. Farrelly, O. M. Johannessen, H. A. Johannessen, W. J. Campbell, P. Gloersen, D. J. Cavalieri, and C. Matzler, "Norwegian remote sensing experiment: Evaluation of the Nimbus-7 scanning multichannel microwave radiometer for sea ice research", *J. Geophys. Res.*, Vol.88, No.C5, pp. 2781-2791, 1983.
- Taylor, V.R., and L.L. Stowe, "Reflectance characteristics of uniform earth and cloud surfaces derived from Nimbus 7 ERB", *J. Geophys. Res.*, Vol.89, No.D4, pp. 4987-4996, 1984.

ACKNOWLEDGEMENTS

This research was funded under contract NAG 5-882 by NASA Ocean Sciences Branch. Additional support is provided by the Swiss Federal Institute of Technology.

SENSITIVITY OF PASSIVE MICROWAVE SEA ICE CONCENTRATION ALGORITHMS TO THE SELECTION OF LOCALLY AND SEASONALLY ADJUSTED TIE POINTS

Konrad Steffen, Swiss Federal Institute of Technology,
Dept. of Geography, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
Email: K.Steffen@omnet(TELEMAIL); Geograph@czheth5a(BITNET)

Axel Schweiger, Cooperative Institute for Research in Environmental Sciences,
University of Colorado at Boulder, Campus Box 449, Boulder, CO, 80303,
Email: Schweiger@Colorado(BITNET); Kryos::Schweiger(SPAN)

ABSTRACT

The sensitivity of passive microwave sea ice concentration algorithms to the selection of tie points was analyzed. Ice concentrations were derived with the NASA Team ice algorithm for global tie points and for locally and seasonally adjusted tie points. The passive microwave ice concentration from the Special Sensor Microwave Imager (SSM/I) were then compared to Landsat MSS derived ice concentrations. Preliminary results show a mean difference of SSM/I and Landsat derived ice concentrations for 50 x 50 km grid cells of 2.7% along the ice edge of the Beaufort Sea during fall with local tie points. The accuracy decreased to 9.7% when global tie points were used. During freeze up in the Beaufort Sea, with grey ice and nilas as dominant ice cover, the mean difference was 4.3% for local tie points and 13.9% for global tie points. For the spring ice cover in the Bering Sea, a mean difference of 4.4% for local tie points and 15.7% for global tie points was found. This large difference reveals some limitations of the NASA-Team algorithm under freeze-up and spring conditions (thin ice areas). In the Weddell Sea of the Antarctica, global tie points perform quite well as the mean difference between Landsat and SSM/I derived ice concentrations was only 3.9%, compared to 2.1% for local tie points.

This analysis indicates that the accuracy of ice concentration calculation based on passive microwave data could be greatly improved when varying sea ice properties are accounted for by the selection of locally and seasonally adjusted tie points.

INTRODUCTION

Long-term monitoring of sea ice is of prime interest for studies of global climate. Polar oceans and their ice cover are more than passive indicators of change in global climate. It is thus of great importance to determine the accuracy with which sea ice parameters such as ice concentration and ice extent can be monitored with today's satellite technology.

On June 19, 1987 the Defense Meteorological Satellite Program (DMSP) launched the special sensor microwave imager (SSM/I). The SSM/I is a passive microwave radiometer which provides near real-time data for operational use and for specific research topic (e.g. sea ice). This work is part of the NASA sponsored sea ice validation plan for the DMSP SSM/I. The objective of this validation plan is to demonstrate a quantitative relationship between the SSM/I-derived sea ice parameters and those parameters derived from other data sets including visible and infrared satellite imagery, aerial photographic and high-resolution microwave imagery from aircraft. The general question to be addressed is with what accuracy (relative to these other observations) can we determine the following sea ice parameters:

- (1) position of the sea ice boundary
- (2) total sea ice concentration
- (3) multi-year sea ice concentration

This paper describes the sensitivity of passive microwave sea ice concentration algorithms to the selection of locally and seasonally adjusted tie points. Landsat MSS images were used for the validation of the SSM/I derived ice concentrations. The technique of ice concentration retrieval from Landsat MSS imagery is described by Schweiger and Steffen (this issue).

ICE ALGORITHM AND TIE POINTS

The derivation of ice concentrations from the SSM/I data was carried out using the algorithm described by Cavalieri et al. (1984), with the addition of the weather filter as discussed by Gloersen and Cavalieri (1986).

The measured brightness temperatures of the 19 GHz vertical and horizontal as well as the 37 GHz vertical channels are interpreted as the respective contributions of open ocean, first-year and multi-year ice.

$$TB = TBW(1-Cf-Cm) + TB_F \cdot Cf + TB_M \cdot Cm \quad (1)$$

where:

- TBW = Brightness temperature of open water
- TB_F = Brightness temperature of first-year ice
- TB_M = Brightness temperature of multi-year ice
- Cf = First-year ice concentration
- Cm = Multi-year ice concentration

The respective concentrations of multi- and first-year ice can thus be expressed as a function of the polarization and gradient ratios and constant tie points for open water, 100 % first-year and 100% multi-year ice.

$$CM = M0 + M1PR + M2GR + M3(PR)(GR) \quad (2)$$

$$CF = F0 + F1PR + F2PR + F3(GR)(PR) \quad (3)$$

$$PR(f) = (TB_V(f) - TB_H(f)) / (TB_V(f) + TB_H(f)) \quad (4)$$

$$GR = (TB_V(37) - TB_V(19)) / (TB_V(37) + TB_V(19)) \quad (5)$$

where TB_V and TB_H are the observed vertically polarized and horizontally polarized brightness temperatures, respectively, and (f) is either 19 or 37 GHz; and $M0, M1, M2, M3, F1, F2, F3$ are constants solely dependent of the tie points. A graphical depiction of the algorithm is shown in Figure 1 with polarization and gradient values plotted for the Beaufort Sea area (Sept. 17, 1987, first case study).

Tie points TBW , TB_F , and TB_M are radiances chosen as typical for open ocean, first-year ice, and multi-year ice derived from a global data set as a mean over one year. However, brightness temperatures of these surfaces are not constant over time because of changes in ice thickness, melt/freeze cycles, snow cover variation, ice roughness changes and drainage of brine pockets. There are large variations in TBW along the ice edge, and in open and very open pack ice areas, caused by various effects including surface roughness, foam and atmospheric water vapor content. A combination of these tends to increase the brightness temperature of the ocean by as much as 40 degree K. Further, the physical properties of ice in different geographical regions can be diverse (e.g. large areas of pancake ice are common in the Antarctica whereas in the Arctic this ice type is only found in the marginal ice zone and in polynyas). Also TB_F are affected by spatial and temporal variations in physical temperature (Steffen and Maslanik, 1988). If globally chosen tie points are used for regional ice concentration calculation and multi-year ice fraction, the variation of TBW , TB_F and TB_M over time and space are ignored.

A preliminary set of global tie points provided by the SSM/I validation team manager (D. Cavalieri) was used to calculate ice concentrations. Local tie points were derived from selected case studies using the following procedure. At least ten 25x25 km grid cells of open water were averaged near the ice edge over one

to three days and used as water tie point (T_{BW}). The same method was applied to select first-year tie points (T_{BP}) and multi-year tie points (T_{BM}) (Table 1). In order to locate areas of first-year ice and multi-year ice, some knowledge of local ice type distribution is essential. For six case studies the ice concentrations were calculated based on globally chosen tie points and on locally chosen tie points. The values were then compared against the ice concentrations derived from the Landsat MSS data (Table 2).

RESULTS

Beaufort Sea

Preliminary results show a mean difference of SSM/I and Landsat derived ice concentrations for 50x50 km grid cells of 2.7 % along the ice edge of the Beaufort Sea during fall (Sept. 17, 1987) with local tie points. The difference increases to 9.7 % with global tie points (Fig. 2). Ice concentrations derived with global tie points were 5 % or more higher than the corresponding ice concentrations derived with local tie points.

During the freeze up in the Beaufort Sea (Nov. 10, 1987), with grey ice and nilas as dominant ice types, the mean difference for the 50 km grid was 4.3 % for local tie points, and 13.9 % for global tie points (Fig. 3). Ice concentrations derived with global tie points were 10 % or more lower than the corresponding ice concentrations derived with local tie points.

For the spring ice cover in the Beaufort Sea (March 12, 1988) the mean difference for 50 km grid cells in Landsat and SSM/I ice concentrations was only 0.8 % with the locally chosen tie points. The same comparison with globally chosen tie points showed a mean difference of 6.9 % with values between -4 % and -10 % (Fig. 4).

The last case study for the Beaufort Sea (March 16, 1988) showed only small differences between Landsat and passive microwave ice concentrations when local tie points were used, whereas ice concentrations derived with global tie points were more than 10 % lower than the corresponding Landsat ice concentrations (Fig. 5).

Bering Sea

For the spring ice cover in the Bering Sea (March 13, 1988), a mean difference of 4.4 % for local tie points, and of 15.7 % for global tie points was found for SSM/I and Landsat ice concentrations (Fig. 6). This large difference - even with locally chosen tie points - reveals some limitations of the NASA-Team algorithm under freeze-up condition. Polarization ratios of new ice and nilas range between 0.05 and 0.2 for 19 GHz (Steffen and Maslanik, 1988). Therefore, the NASA Team algorithm underestimates ice concentrations when new ice and nilas are covering large areas.

Weddell Sea

In the Weddell Sea of the Antarctica (Nov. 29, 1987), global tie points perform quite well as the mean difference between Landsat and SSM/I derived ice concentrations was only 3.9 %, compared to 2.1 % for local tie points (Fig. 7).

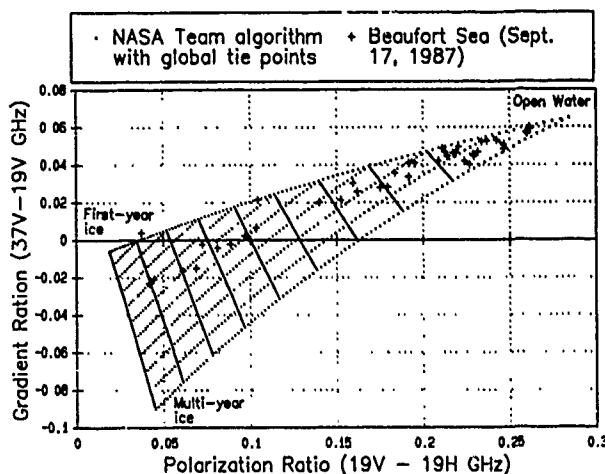


Fig. 1 Graphical depiction of SSM/I algorithm. The inside of the triangle describes total ice concentrations from 0 to 100 % along the longer sides of the triangle and multi-year ice fractions along the shorter side. The corners of the triangle represent tie points for open water, 100 % first-year ice, and 100 % multi-year ice (After Cavalieri and Gleresen, 1986). The crosses give the polarization and gradient ratios for case study 1 (Beaufort Sea, Sept. 17, 1987).

| Case | Mean | Std Dev | Min | Max | # |
|------|------|---------|-----|------|----|
| 1 | 2.7 | 2.7 | 0.2 | 5.9 | 9 |
| 2 | 4.3 | 2.9 | 1.6 | 9.7 | 6 |
| 3 | 0.8 | 0.7 | 0.1 | 1.7 | 7 |
| 4 | 1.5 | 1.3 | 0.2 | 3.8 | 12 |
| 5 | 4.6 | 4.1 | 0.0 | 13.1 | 14 |
| 6 | 2.1 | 2.1 | 0.0 | 5.7 | 7 |

Number of grid cells (50x50 km)

Table 2 Differences between Landsat ice concentration and SSM/I ice concentration based on locally chosen tie points. The Min and Max represent the smallest respectively largest difference in ice concentration for an individual grid cell. Cases 1 to 4 are from the Beaufort Sea corresponding to Figs. 2 to 5, case 5 is from the Bering Sea (Fig. 6), and case 6 from the Weddell Sea (Fig. 7).

| Channel | NASA | 1 | 2 | 3 | 4 | 5 | 6 |
|---------|-------|-----|-----|-----|-----|-----|-----|
| 19V OW | 176.3 | 187 | 185 | 177 | 182 | 184 | 177 |
| 19V FY | 252.4 | 245 | 252 | 258 | 254 | 257 | 266 |
| 19V MY | 221.8 | 222 | 222 | 228 | 218 | 222 | 222 |
| 19H OW | 96.1 | 120 | 115 | 100 | 110 | 102 | 103 |
| 19H FY | 244.8 | 230 | 232 | 241 | 238 | 238 | 254 |
| 19H MY | 202.3 | 202 | 202 | 204 | 198 | 202 | 202 |
| 37V OW | 201.7 | 209 | 205 | 200 | 204 | 204 | 187 |
| 37V FY | 243.6 | 245 | 250 | 255 | 250 | 254 | 261 |
| 37V MY | 183.5 | 184 | 184 | 196 | 188 | 184 | 184 |

Table 1 Tie points used in the sensitivity study of passive microwave sea ice concentration algorithm to the selection of globally (NASA) and locally adjusted values (case studies 1 to 6, see Figs. 2-7). Tie points are given in degree Kelvin for open water (OW), first-year ice (FY), and multi-year ice (MY), and for the frequencies 19 GHz vertical and horizontal polarizations, and 37 GHz vertical polarization.

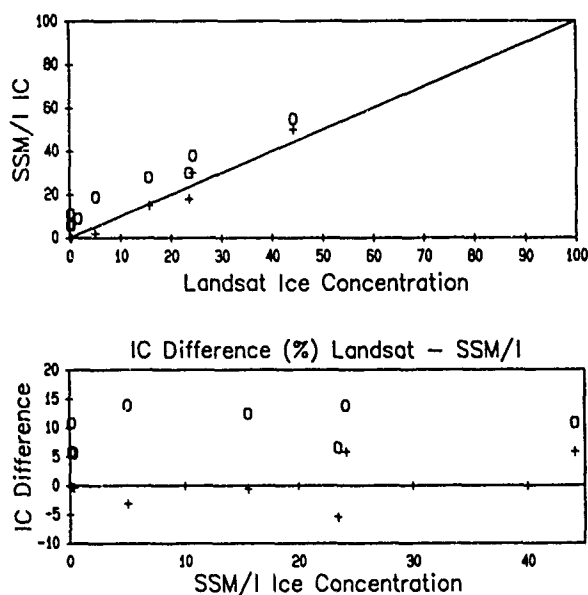


Fig. 2 Comparison of ice concentration (50 km grid) between NASA-Team algorithm using local, respectively global tie points versus Landsat values for the Beaufort Sea (Sept. 17, 1987). Globally chosen tie points are called NASA tie points in graph.

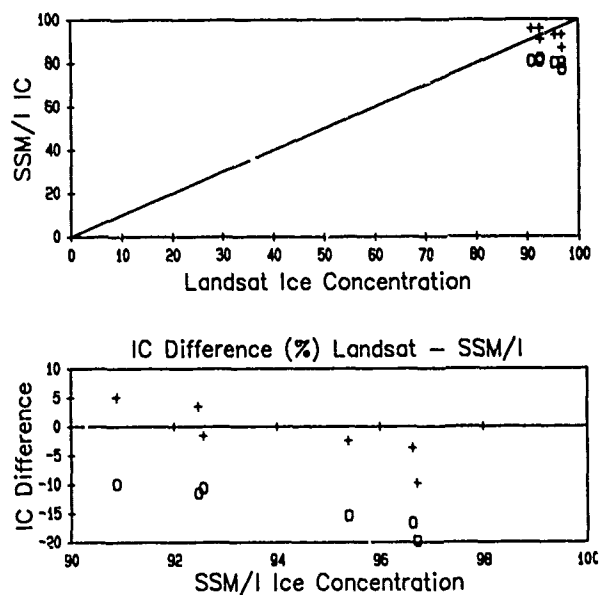


Fig. 3 Same then Fig. 2 for Beaufort Sea (Nov. 10, 1987).

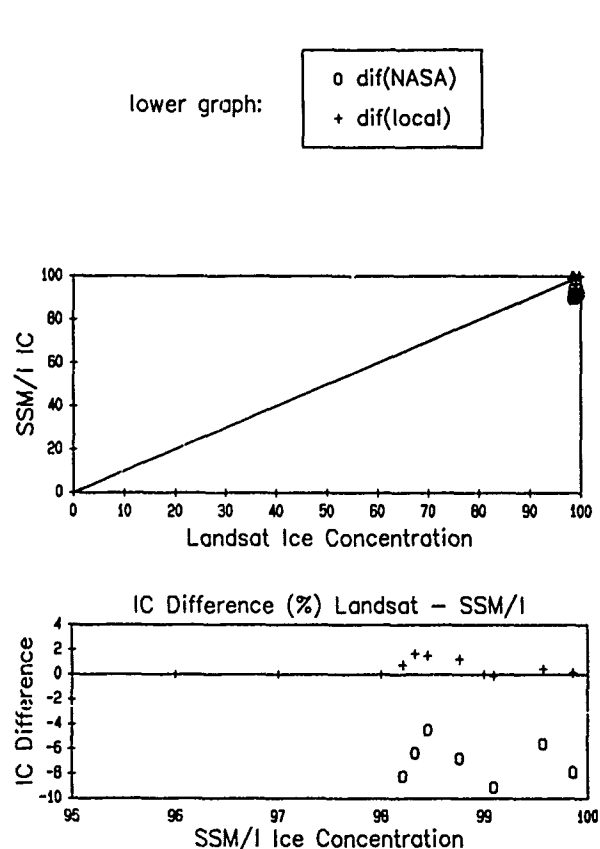


Fig. 4 Same then Fig. 2 for Beaufort Sea (March 12, 1988).

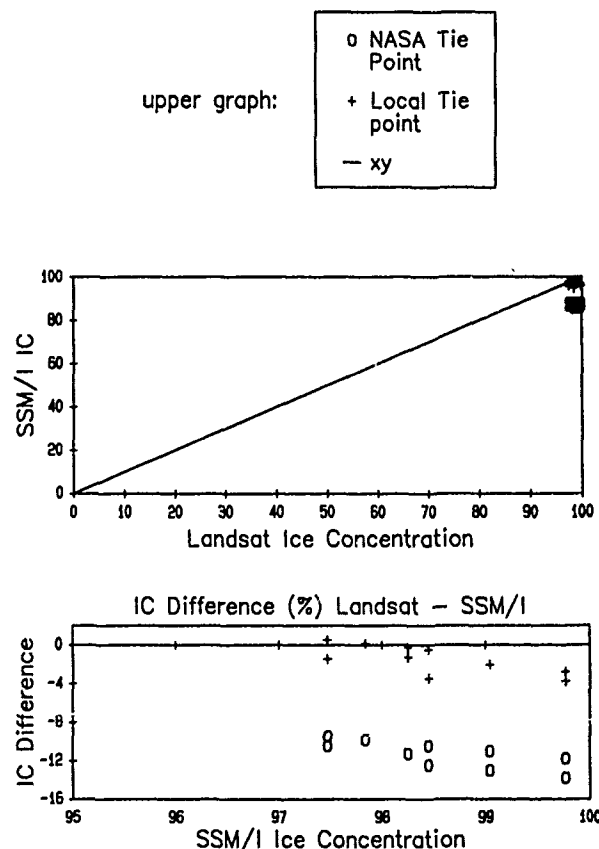


Fig. 5 Same then Fig. 2 for Beaufort Sea (March 16, 1988).

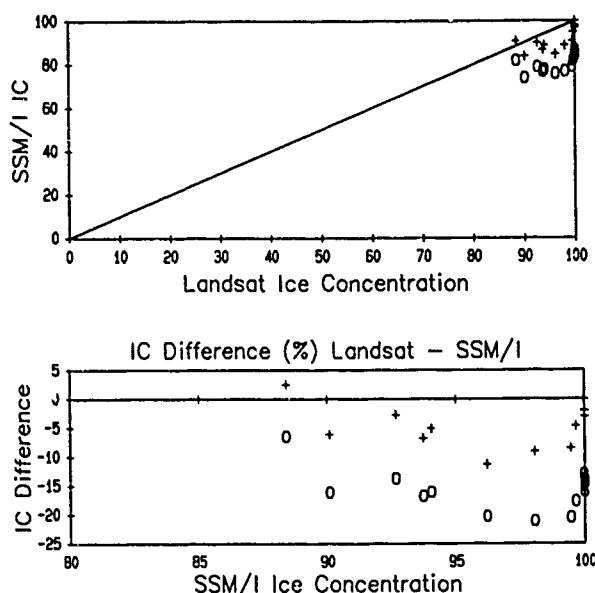


Fig. 6 Same then Fig. 2 for Bering Sea (March 13, 1988).

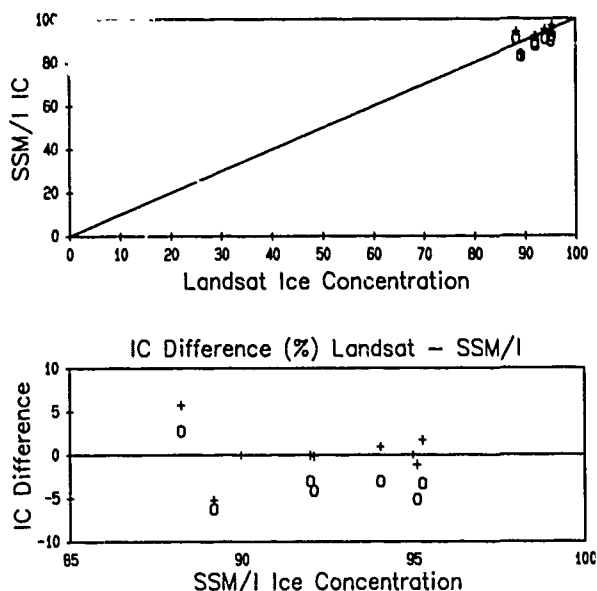


Fig. 7 Same then Fig. 2 for Weddell Sea (Nov. 29, 1987).

lower graph:

o dif(NASA)
+ dif(local)

upper graph:

o NASA Tie
Point
+ Local Tie
point
— xy

CONCLUSION

In areas with higher amounts of nilas and young ice we found that the SSM/I ice concentration algorithm underestimates ice concentration by as much as 11%. This was observed in two separate Landsat-SSM/I comparisons (Beaufort Sea: Nov. 10, 1987, and March 13, 1988). Due to the higher polarization ratio of nilas (ice types which are less than 0.1 m in thickness) compared to white ice at 19 GHz, the NASA Team algorithm underestimates ice concentration considerably.

It appears that seasonal and regional adjusted tie points will improve the overall performance of the SSM/I sea ice concentration algorithms, at least in the North Polar region. Our work suggests that where the ice has a higher variation internal characteristics (e.g. salinity) that the global tie points will cause a mean error of up to 10% in the ice concentration. Use of local tie points dropped the mean error to 3% for the same cases studied. In the South Polar regions, where the ice cover is essentially one year in age, the global tie points offer reasonable performance. The mean error was only decreased from 4% for global to 2% for local tie points.

ACKNOWLEDGEMENTS

This research was funded under contract NAG 5-882 by NASA Ocean Science Branch. Additional support is provided by the Swiss Federal Institute of Technology. Thanks are due to Ron Weaver and Roger Barry of NSIDC for their scientific inputs as well as administrative assistance. Don Cavalieri's suggestions as well as Bob Thomas's encouragement were invaluable in this research. Charles Morris at NODS and Frank Wentz at Remote Sensing Systems provided SSM/I data.

REFERENCES

- Cavalieri, D. J., P. Gloersen, and W. J. Campbell, "Determination of sea ice parameters with Nimbus 7 SMMR", *J. Geophys. Res.*, Vol. 89, pp. 5355-5369, 1984.
- Gloersen P., and D.J. Cavalieri, "Reduction of weather effects in the calculation of sea ice concentration from microwave radiances", *J. Geophys. Res.*, Vol. 91, No. C3, pp. 3913-3919, 1986.
- Steffen, K., and J.A. Maslanik, "Comparison of Nimbus 7 Scanning Multichannel Microwave Radiometer radiance derived from ice concentrations with Landsat imagery for the North Water area of Baffin Bay", *J. Geophys. Res.*, Vol. 93, No. C9, pp. 10769-10781, 1988.

UTILIZATION OF LOCAL TEXTURE TRANSFORMS FOR ADAPTIVE FILTERING OF SAR SEA ICE IMAGERY

David F. Salter and Ellsworth F. LeDrew

Department of Geography
University of Waterloo
Waterloo, Ontario
N2L 3G1

Abstract

In this paper we report on the utilization of texture measures generated from Grey Level Co-occurrence Matrix (GLCM) as the basis for filtering Synthetic Aperture Radar sea ice imagery. Four measures of texture were generated and then processed through an adaptive local statistics filter. The filtered images were subsequently algebraically summed to the original image to provide a viable and interpretable tonal product. Results show that the filtered texture measure "Contrast/Original" provides good noise reduction and ice ridge retention.

Keywords: Texture measure, Adaptive filtering, Synthetic Aperture Radar, Sea ice.

1. Introduction

Increasing economic development of the Canadian Arctic has created demand for an extended shipping season and therefore improved ice navigation ability. Utilization of real-time Synthetic Aperture Radar (SAR) imagery permits long range and strategic route selection through ice covered waters.

SAR imagery, due to its multiplicative nature, is degraded by random field speckle. Because speckle obeys a negative exponential behaviour, the noise is concentrated in the same tonal regions as physically significant pressured ice features, such as ridges, which are characterised by bright corner reflected and Lambertian returns. The presence of speckle may introduce distortions in feature morphology or mask the feature altogether.

Techniques developed to reduce image speckle while maintaining the integrity of feature morphology have focused on the use of adaptive spectral image filters based on local statistic transforms [1-8]. Lee [5] developed a series of filters based on the assumption that the mean and variance of a sample pixel are equal to the local mean and variance of its surrounding neighbours. The premise is that any

pixel values within the defined subarea and lying outside of the 2-sigma range represent either noise or edge detail. A gradient threshold factor was employed to identify the presence of an edge within this region.

The disadvantage of these filters is that only the spectral properties of the image are considered. Therefore the next logical direction for examining image filtering is the utilization of image texture. There has been extensive research into the use of texture in remotely sensed imagery for the purpose of improving ice classification. However, the majority of these studies have not investigated the application of texture for input to local adaptive filters. In this paper we report on the use of local texture transforms as a means of filtering SAR sea ice imagery for the purpose of both image noise reduction and pressure ridge detection and enhancement.

2. Study Area

The study area in the North Wellington Channel, lat. 76° N/long. 98° W imaged Dec. 2, 1986, has been selected on the basis of its content of definitive ridges and large areas of homogeneous smooth First-Year Ice (FYI). The imagery is Intera STAR-1 X-band, three line averaged and azimuthally resampled to 12m² resolution provided through the co-operative arrangement with Canarctic Shipping Company Ltd.

The sub area of 256x256 pixels was selected from the test scene and reduced to 32 grey levels to facilitate computation [12]. Although the use of the small image size may preclude some coarser textural variations, the image content is considered to be sufficiently varied for this evaluation.

3. Texture Filter

The texture measures used were based on the Spatial Grey Level Co-occurrence Dependence Matrix (GLCM) developed by Haralick et al [9-10]. The GLCM is based on the assumption that the average spatial relationship of grey tones in a local image window are contained in the average spatial relationship of grey tones in the image. This is the same assumption underlying spectral local adaptive filters [6]. Therefore, texture measures derived from local texture transforms do not require *a priori* knowledge of the image model.

The GLCM generated was based on the estimation of second order joint conditional probability density function, where $P(i,j|d,\theta)$ is the probability of moving from grey level i to grey level j , given the polar form intersample spacing distance d (where $d = \max[\Delta x, \Delta y]$) and angular orientation θ (where $\theta = \arctan(\Delta x/\Delta y)$) [10]. The orientations selected were 0° , 45° , 90° and 135° , where $\Phi(d,\theta) = \Phi^T(d,\theta)$, its transpose matrix [11].

The four orientations were averaged to provide the directional measure. The intersample distance (d) was set to 1 in accordance with Haralicks' findings on information majority content [9].

GLCM's were generated for window sizes of 17×17 and 21×21 .

The texture measures employed were based on a qualitative analysis of the contextual information of the images. The presence of large homogeneous FYI regions and distinct pressure ridges lend best to description by texture measures which correspond to image homogeneity, contrast and correlation. Consequently, the texture measures of 'Uniformity', 'Contrast', 'Inverse Difference Moment' and 'Entropy' were selected.

$$\begin{aligned} \text{Uniformity} & \sum_{i=1}^n \sum_{j=1}^n P_{ij}^2 \\ \text{Contrast} & \sum_{i=1}^n \sum_{j=1}^n P_{ij} (i-j)^2 \\ \text{Inverse Difference Moment} & \sum_{i=1}^n \sum_{j=1}^n \frac{P_{ij}}{1+|i-j|} \\ \text{Entropy} & \sum_{i=1}^n \sum_{j=1}^n P_{ij} \log P_{ij} \end{aligned}$$

where

P_{ij} is the GLCM matrix value at position (i,j) .
 n is the number of pixel pairs in the image at (α, δ) ,
 σ_x is the sample standard deviation of row i ,
 σ_y is the sample standard deviation of row j ,
 μ_x is the sample mean of row i ,
 μ_y is the sample mean of row j .

A 17×17 moving window was applied to the 32 grey level image with the output central pixel written to an interim image which formed the base for the texture filtering operation.

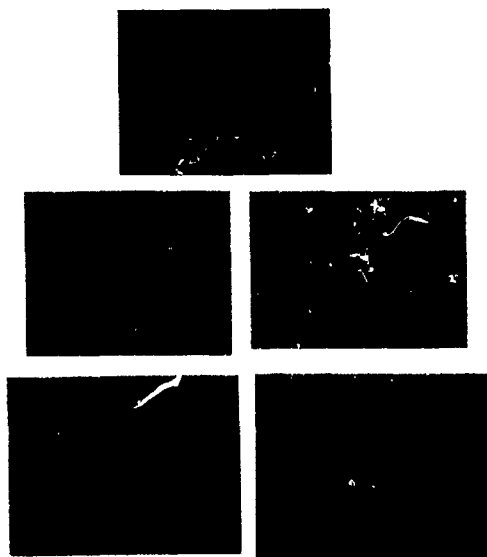
An adaptive 3×3 moving window based on a variation of the Lee sigma filter [5] was applied to the texture image. Pixels with values outside of the 2-sigma range were excluded from further processing. The remaining pixels, if satisfying the 5 majority rule, were converged to create the new output pixel, otherwise the new output pixel was the result of the mean window value.

In order to compare results the images were normalised and stretched to 256 grey levels. The filtered texture images on their own did not provide a suitable display. To this end the filtered texture measure images were algebraically ratioed to the original image through a difference over sum operation: $T = F(1) - F(2) / F(1) + F(2)$ where $F(1)$ is the original image and $F(2)$ is the filtered texture image. This allowed for direct comparison and ordinal ranking.

4. Results and Discussion

Preliminary results indicate that the texture filter is a viable tool for image noise reduction and edge retention in SAR sea ice imagery. Qualitatively, the filtered texture images clearly reduced the noise in the homogeneous FYI and retained the ridges with varying degrees of success (figures 1-5). The Uniformity/Original (UO) image lost all low order ridges while distorting the areal extent of the higher order ridges, where ridge order is a function of pixel width (ie. the greater the width the higher the ridge order). The FYI became very uniform in contrast and tone. The result was a low contrast image with sparse detail regarding the ice content. The Contrast/Original (CO) image provided greater delineation of the ridges while retaining some natural variation in the FYI. The Inverse Difference Moment/Original (IO) image maintained ridge integrity and smoothed the FYI while retaining some natural variation. Ridges in the Entropy/Original (EO) image were reduced in tonal value and subsequently suffered in ridge/FYI ratio. The FYI was smoothed to the extent of natural variation removal.

The images were quantitatively compared through image contrast (μ/σ), ridge to FYI ratio, and ridge integrity (width) retention measures. The latter is a measure of ridge pixel difference between the original and filtered ridges. The summary statistics are presented in Table 11.



Figures (1) Original image, lat. 76° N/long 98° W, Dec. 2, 1986, SAR X-band 12m² (2) Uniformity/Original, (3) Contrast/Original, (4) Inverse Difference Moment/Original, and (5) Entropy/Original images.

Table 1. Evaluation of Texture filter performance

| Image | Contrast (μ/σ) | Ratio | Ridge Retention |
|----------|---------------------------|-------|-----------------|
| Original | 0.5023 | 1.276 | 1, 5, 6 |
| UO | 0.5128 | 1.056 | 0, 5, 6 |
| CO | 0.5128 | 1.215 | 1, 6, 6 |
| IO | 0.5126 | 1.351 | 1, 6, 7 |
| EO | 0.5128 | 1.450 | 0, 7, 8 |

¹ Based on average of 5 selected test sites.

From Table 1 it can be seen the Contrast/Original (CO) image provides the best overall results. The ridge integrity was largely maintained, with a slight spatial extension of the 5 pixel wide ridge. The ratio of ridge to FYI most closely approximated the original image and provided the best discrimination. These results correlate with the qualitative analysis as the CO visually provided the best filtered representation of the original image. The Inverse Moment Difference (IO) provided the next best filtered image both qualitatively and statistically. The Uniformity (UO) image was the poorest performer on the basis of all parameters tested.

5. Conclusions

Our analysis has shown that on a simple ridge/FYI image the texture filter is a viable tool for noise reduction and ridge retention. From the measures tested the combination of algebraically summed Contrast and original image provided the best filtering results while the Uniformity and original image combination was least effective.

Future work will involve the use of other textural measures suited to more heterogeneous images. As this study was limited to the use of a single window size, further testing will be conducted into the effect of window size and image content retention. As well new techniques for presenting the filtered texture images will be explored.

Acknowledgements

This research was supported by an NSERC operating grant to E. LeDrew and an NSERC strategic grant to Drs. Jernigan, Sykes, LeDrew and Howarth. The authors would also thank Canarctic Shipping Company Ltd. for providing the Intera STAR-1 SAR digital data.

References Cited

1. Lee, J.S., "Digital image enhancement and noise filtering by use of local statistics". IEEE Trans. Pat. Analysis and Mach. Intel. Vol. 2, No.2, pp165-168, 1980.
2. Lee, J.S., "Refined filtering of image noise using local statistics". Comp.Vis. Graph. and Img. Proc. Vol.15, pp380-389, 1981a.
3. Lee, J.S., "Speckle analysis and smoothing of synthetic aperture radar." Comp. Graph. and Img. Proc. Vol.17, pp24-32, 1981b.
4. Lee, J.S., "A simple speckle smoothing algorithm for synthetic aperture radar imagery." IEEE Trans. on Sys., Man and Cyber. Vol.13, No.1, pp85-89, 1983.
5. Lee, J.S., "Digital image filtering and the sigma filter." Comp. Vis., Graph. and Img.Proc. Vol. 24, No.2, pp 255-269, 1983.
6. Frost, V.S., J.A. Stiles, K.S. Shanmugan, J.C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise." IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.4, No.2, pp157-165, 1982.
7. Frost, V.S., M.S. Perry, et al. 1983. "Digital enhancement of SAR imagery as an aid in geologic extraction." P. Eng. & R. S., Vol. 49, No.3, pp357-364, 1982.
8. Kuan, D.T., A.A. Sawchuk, T.C. Strand, P. Chavel., "Adaptive noise smoothing filter for images with signal-dependent noise." IEEE Trans. Pat. Rec., Mach.Intel. Vol.7, No. 2, pp165-177, 1985.
9. Haralick, R.M. K., Shanmugan, I. Dinstein. "Textural features for Image classification." IEEE Trans. Sys.Man Cyber. Vol.3, No.6, pp165-177, 1973.
10. Haralick, R.M., "Edge region analysis for digital image data." Comp. Graph., Vis. and Img. Proc. Vol.12, pp60-73, 1980.
11. Connors, R.W. and C.A. Harloe, "A theoretical comparison of texture algorithms." IEEE Trans. Pat. Analysis Mach. Intel. Vol.2, No.3, pp204-222, 1980.
12. Shangmugan K.S., V. Narayanam, V.S. Frost, J.C. Holtzman, "Textural Features for Radar Image Analysis" IEEE Trans. Geo and R. S. Vol 19, No.3, pp153-156, 1981.

SPRING SURFACE CIRCULATION PATTERNS DETECTED USING REMOTE SENSING OF DRIFTING ICE FLOES IN HUDSON BAY, CANADA

Pierre Larouche

Institut Maurice-Lamontagne, Ministère des Pêches et des Océans,
C. P. 1000, Mont-Joli, Québec, Canada G5H 3Z4
Tel: (418)-775-6569, Telex: 051-6303, Fax: (418)-775-6542

ABSTRACT

Evaluation of surface currents using remote sensing (Landsat MSS) of free drifting floes and a dynamical approach lead to the recognition of two modes of circulation in southeastern Hudson Bay in the spring of 1985. Of the different factors that can generate these two modes, the wind seems the most highly related to the observed patterns. The circulation induced by strong persistent winds is also shown to subsist for at least three days after the end of the storm.

Keywords: Ice drift, Circulation, Remote sensing, Wind

1. INTRODUCTION

Evaluating surface circulation in ice infested areas is one of the most difficult task to accomplish in physical oceanography. Over the years, the use of the Lagrangian approach using drifting ice floes as tracers of the water movement proved much helpful in the determination of the surface circulation in arctic areas. There is however another force that influence ice drift: the wind. It has been showed that wind is the main driving force over time scales ranging from days to months (Thorndike and Colony, 1982). So the drift of ice for these time scales may very well be more representative of the wind pattern than of the surface circulation. It does not however mean that the water circulation is negligible for ice drift but simply that it may be concealed and more difficult to evaluate.

It has recently been showed that the use of a dynamical approach to calculate the surface currents underlying drifting floes produces better results than measuring ice drift only (Larouche, 1989). This methodology is thus used in order to evaluate the spring circulation pattern in southeastern Hudson Bay (Fig. 1), an area where many winter field surveys were done over the last 7 years. These data should help complement our understanding of the general circulation patterns in this area and of the factors affecting it.

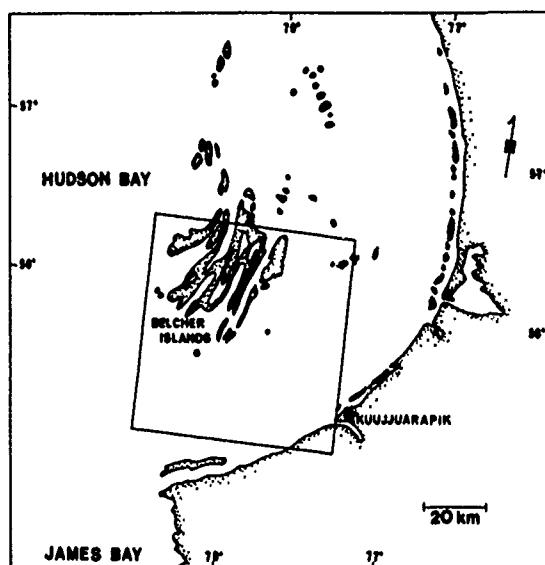


Fig. 1 Map of the area indicating the location (square) of the evaluated currents on fig. 2.

2. METHODOLOGY

The methodology used is quite simple. It is based on the free drift equation relating the three main forces acting on ice drift: wind, current and Coriolis force. First, ice drift is evaluated relative to selected landmarks using sequential satellite pictures. These data are then input, together with wind data for the area, in a software that model the ice drift using the wind and an approximate value of the current. The current is then adjusted until the difference between the modeled and observed drift is less than 50 meters. The result is a good approximation of the tidally averaged current as the time between the satellite pictures (~24 h) is very close to the diurnal tidal period (24.8 h).

In this survey, we use 4 pairs of black and white prints of Landsat MSS-4 (near IR) images taken in the spring 1985 between the beginning of May and mid-June.

This choice of satellite is mainly made on the basis of the better ground resolution of Landsat (80 m) as compare to NOAA images (1,1 km) thus allowing the identification of much smaller floes. All images show a number of free drifting floes over the area allowing us to evaluate the circulation pattern. Positions of recognizable features on the perimeter of the floes are directly digitized from the black and white prints at a scale of 1 : 1 000 000. These data are then corrected for scale differences between the images using the known distance between the visible landmarks on the images.

To drive the model, geostrophic wind climatology (GWC) data coming from the Atmospheric Environment Service of Canada is used. This choice of wind data set is made because of their better correlation with winds evaluated from surface pressure maps (Olson, 1986) for this particular portion of Hudson Bay. Directions were rotated 20° to the right as recommended by Swail et al. (1984).

3. RESULTS

Figure 2 shows the resulting currents evaluated for the four pairs of images. We can see that the surface circulation pattern seems variable with time showing two principal modes. In the first two weeks of May, the first mode indicates a circulation directed toward the SW. The inverse situation (mode 2) is seen in early June when water is flowing into the area. The last image pairs taken in mid-June show a more intricate pattern with water flowing SW along the Belcher Islands coast and indications of a water inflow in the southern portion of the area. This suggests an intermediate situation between the two principal modes.

Only a few factors can influence surface circulation. These are the wind and atmospheric pressure systems, the low frequency tidal signal and influence from other water bodies (James Bay). We will now look at each of these factors.

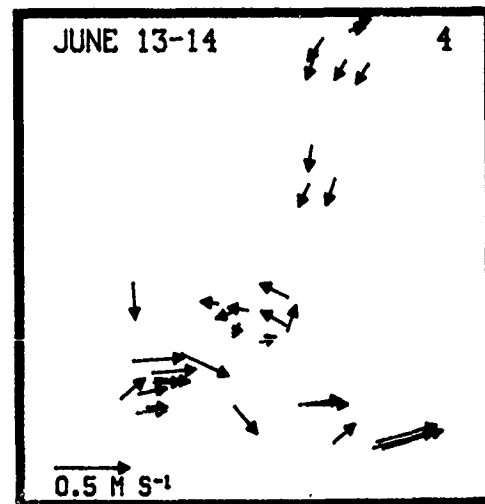
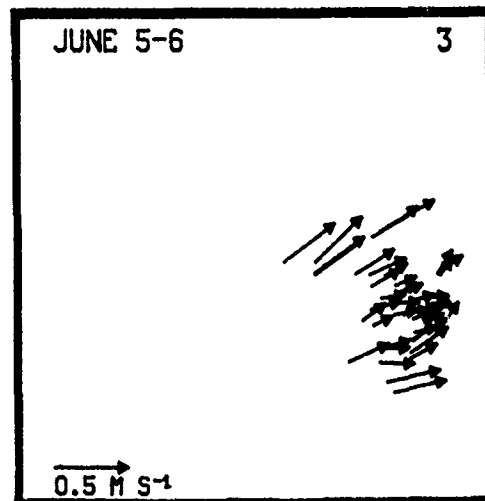
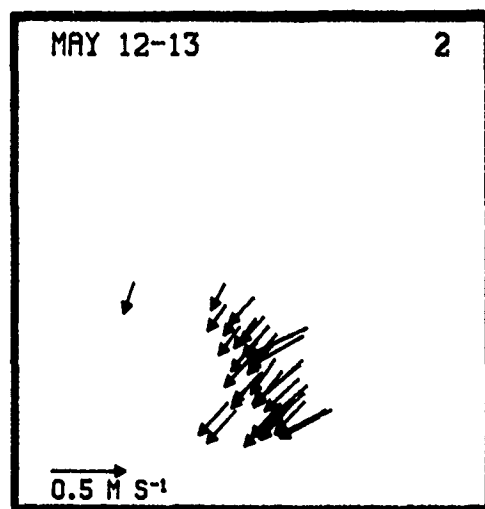
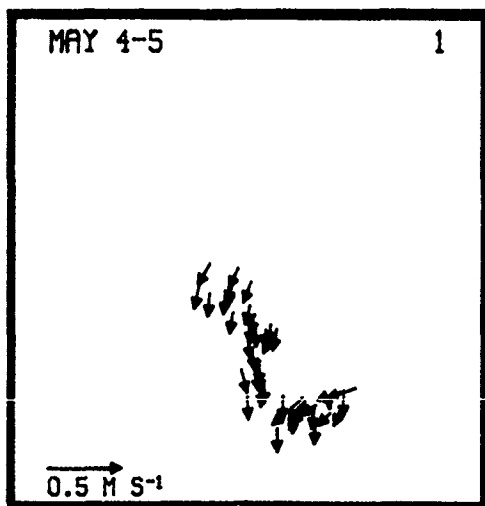


Fig. 2. Surface circulation evaluated for the four pairs of images (1 - 4). See figure 1 for the geographic location.

3.1 Long distance influence

The survey area is located to the east of the Hudson Bay - James Bay junction and as such may be influenced by processes originating in James Bay. The biggest influence from James Bay at this time of year is the freshwater pulse coming from its numerous rivers. This pulse generates a density driven coastal current that propagates northwesterly along the coast until it meets with the cyclonic circulation of Hudson Bay (Prinsenber, 1982a) that could divert it into our survey area. The spring swelling in James Bay normally takes place in mid-April. This is 7-8 weeks before the first sign of water input into our area is seen. Using an average current speed of 0.1 m s^{-1} (Prinsenber, 1982b), this leads to a travelling distance of 430 km. This is however much higher than the actual distance between the rivers and our area. So it seems unlikely that the spring freshwater pulse is responsible for the sharp change of circulation pattern seen from the satellite imagery.

3.2 Tidal signal

The second possible cause for the change in circulation pattern is the low frequency (neap-spring) tidal signal. Tides in this area are relatively small ($\sim 1.6 \text{ m}$ in Kuujjuarapik). However, as the water depth is relatively shallow ($\sim 100 \text{ m}$), differences in the spring-neap tides could be reflected in the circulation patterns. Figure 3 shows the predicted tidal signal in Kuujjuarapik over the time the images were taken. No correlation exists between the low frequency tidal signal and the surface circulation patterns as both the first and second modes are observed close to spring tides, the first mode being also observed close to a neap tide. This factor can thus also be neglected as the source of the observed variability.

3.3 Wind and atmospheric pressure systems

Atmospheric pressure systems influence the water circulation in two ways. First, they generate the wind field that have a direct action on the circulation of the surface water. They also have an indirect influence on circulation by generating a sea surface tilt that in turn leads to a current. This current is however much smaller than the one due to the direct wind stress on the sea surface so we can reasonably neglect it in this study.

Direct wind action is thus the only remaining force that can possibly explain the different circulation patterns observed in this study. Sea surface pressure maps available for these dates (not shown) indicates that the surface stress correlates very well both in amplitude and direction with most of the calculated currents. The only exception is the intermediate circulation pattern that does not fit completely with the observed pressure system. A look at the wind record prior to the images acquisition show that the surface circulation seems to have reacted in less than a day to the changing wind for the May 12 and June 5 evaluations.

The difference with the June 12 situation was the duration of the wind stress applied on the sea surface. Winds in Kuujjuarapik have been recorded blowing mainly from the SW, sometimes very strong, for seven days ending June 11. This was followed by a period of relatively light winds from the north until the satellite images were taken two and three days later. On the other hand, the duration of the wind was never more than two days for the other two evaluations. Surface water circulation in this area thus seems to be mainly driven by the wind even over short periods of time. However, the effect of this applied stress can be felt for at least three days when a strong SW impulse is given to the system.

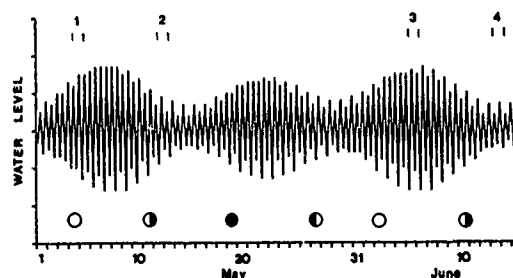


Fig. 3. Predicted tidal signal in Kuujjuarapik for the spring 1985. Numbers refers to fig. 2.

CONCLUSION

The use of a dynamical approach proved helpful in determining the surface circulation pattern and its driving force in southeastern Hudson Bay. The future availability of space based radars will increase the utilization potential of this method as images will become available with a greater time resolution.

REFERENCES

- [1] Larouche, P., "Evaluation dynamique de la circulation de surface par la d rive de blocs de glace (Baie d'Hudson, Canada)", Photo Interpr tation, 1988, In press.
- [2] Olson, R., "An assessment of Canadian arctic wind data sets", Canadian Climate Centre report 86-9, Atmospheric Environment Service, Downsview, Ontario, 58p., "Unpublished manuscript", 1986.
- [3] Prinsenber, S.J., "Time variability of physical oceanographic parameters in Hudson Bay", Le Naturaliste Canadien, Vol. 109, pp 685-700, 1982a.
- [4] Prinsenber, S.J., "Present and future circulation and salinity in James Bay", Le Naturaliste Canadien, Vol. 109, pp 827-841, 1982b.
- [5] Swail, V.R., L.D. Mortsch and D.A. Carr, "Intercomparison of marine winds data sets", Canadian Climate Centre report 34-15, Atmospheric Environment Service, Downsview, Ontario, 84p., "Unpublished manuscript", 1984.
- [6] Thorndike, A.S. and R. Colony, "Sea ice motion in response to geostrophic winds", J. of Geophys. Res., Vol 87, pp 5845-5852, 1982.

DETAILED VEGETATION MAPPING FROM MEIS-II DIGITAL DATA

Paul M. Treitz, Philip J. Howarth, Roger C. Suffling and Paul Smith

Faculty of Environmental Studies, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

Abstract

A study has been undertaken to determine the extent to which high resolution digital data can be used to enhance field methods for vegetation mapping. Data from 155 vegetation plots at Presqu'ile Provincial Park, Ontario are analyzed using TWINSpan cluster analysis to produce several hierarchical ecological classification schemes. These are used to undertake supervised classifications of MEIS II digital data recorded in five spectral bands with a 5 m x 5 m spatial resolution. Accuracy assessments show a considerable range in the classification accuracies for different plant communities. In general, the more ecologically/spectrally unique the class is and the lower the species/spectral variations that occur, the higher will be the classification accuracy.

Keywords: Accuracy assessment, MEIS, Supervised classification, TWINSpan, Vegetation mapping

Introduction

The general capabilities of Landsat MSS and Landsat TM data for vegetation mapping are well documented in the literature (e.g., Hopkins et al., 1988; Talbot and Markon, 1988; Craighead et al., 1988). The level of detail that this imagery provides, however, is relatively low compared with many of the mapping needs of ecologists and foresters. Detailed vegetation mapping usually involves an ecological approach where information, not only on trees but also on shrubs and herbaceous cover is combined into one classification scheme. Using ground plots, vegetation may be classified either subjectively or by using standard multivariate analysis techniques. Phytosociologists have tended to employ multivariate statistical classifications, whereas remote sensing practitioners have usually used subjective classifications. Based on the classification scheme, mapping of the different habitats is carried out, usually in the field with the aid of aerial photographs.

The question arises as to what extent high resolution digital data can be combined with detailed ground information to improve the accuracies and speed of vegetation mapping over relatively large areas. This report forms part of a study on user needs and practices in vegetation mapping. In this paper, results of mapping diverse vegetation from high

resolution airborne digital data acquired by the MEIS II system are presented.

Study Area

The study area is Presqu'ile Provincial Park on the north shore of Lake Ontario. The park is located on a boot-shaped tombola peninsula composed of dune ridges which encloses the sheltered and marshy Presqu'ile Bay (Figure 1). The range of landforms, soil types and hydrologic environments produces great vegetational diversity including both uplands and wetlands. Human interference here has resulted in abrupt spatial changes in vegetation, whereas more gradual and subtle changes are observed in undisturbed habitats.

Data Sources

Airborne Data

Airborne digital data were acquired with the Multi-detector Electro-optical Imaging Scanner (MEIS II) (McColl et al., 1983). The data were recorded on June 12, 1984 at 13:00 GMT, from a flying height of approximately 6,100 m. This provided data with a pixel size of 5 m x 5 m.

The higher the spatial resolution of the data, the less ground area is covered by each swath of data. The 5 m x 5 m pixel was selected to provide reasonable areal coverage at a relatively high spatial resolution.

In terms of spectral resolution, the "forestry" set of filters was selected for the study as it was felt that this would produce maximum differentiation in the vegetation. The "forestry" filter set consists of five narrow (3 nm) spectral bands, the bands being centred at 487 nm (blue), 577 nm (green), 664 nm (red), 769 nm (red/near infrared boundary) and 833 nm (infrared).

Field Data

Training plots and verification plots were selected to be broadly representative of the major cover types. Vegetation information was recorded from 155 plots, each measuring 10 m x 20 m and which were representative of neighbouring vegetation. The long axes of the plots were chosen to run east-west, parallel to the flight lines. In each plot, tree and shrub clump densities and herb frequencies were recorded.



Figure 1. A black-and-white photograph showing part of the Presqu'île study area displayed on a colour monitor. The beach ridges and sand dunes can be seen on the west and south sides of the figure. To the north, a series of vegetated recurved spits enter Presqu'île Bay.

Methodology

Four groups of studies were undertaken, in each case a different classification scheme being used. The methodological procedure, however, was similar for each study (Figure 2).

TWINSpan Analysis

The phytosociological vegetation classification was carried out using TWINSpan cluster analysis (Hill, 1979). The result of the analysis is a dichotomous key which provides ecological groupings of tree, shrub and herbaceous covers to be used in the classification. In this paper, the objective TWINSpan results were trimmed subjectively to establish appropriate classes (Figure 3).

In developing the classification scheme and also when selecting the training plots for the field data, the following guidelines were established:

- In establishing the classification, only classes from a similar part of the TWINSpan hierarchy could be merged.

- Classes in a similar part of the hierarchy but displaying different spectral characteristics in their training plot statistics were not merged.

- Minimum training area size was 50 pixels, but larger sample sizes were preferred.

Selection of Training Plots

Training area definition was performed on a false-colour composite generated from the infrared, red and blue bands of the MEIS data which was displayed on the colour monitor of a Dipix ARIES III image analysis system. From the 155 training plots, 119 were selected for the training of the classifier, while the rest were used for later verification. The detailed ground plots measuring 10 m x 20 m were displayed on maps of the study area and written site descriptions of their locations were also available. The boundaries of the ground plots were transferred to the digital image visually. They were enlarged to a minimum size of 15 m x 15 m, and in some cases larger in order to accumulate a significant number of pixels for statistical analysis. However, enlargement of plots was only performed in areas where the make-up of the site was homogeneous and matched that of the actual ground plot.

Classification

Classification was carried out at two levels of detail. In the first trial, a total of 27 TWINSpan classes was used, while in the second case these classes were recombined to give a total of 20 classes. In this paper, only the results of the first trial are reported. In both cases, the maximum-likelihood classifier was used to generate the results.

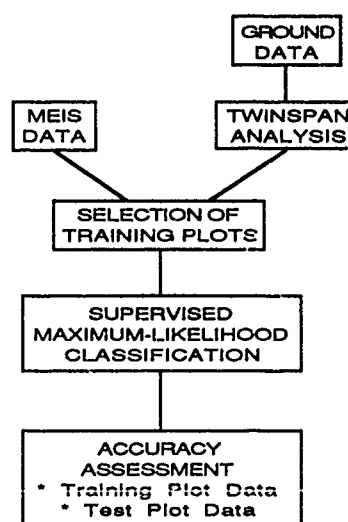


Figure 2. An outline of the methodological procedures followed in this study.

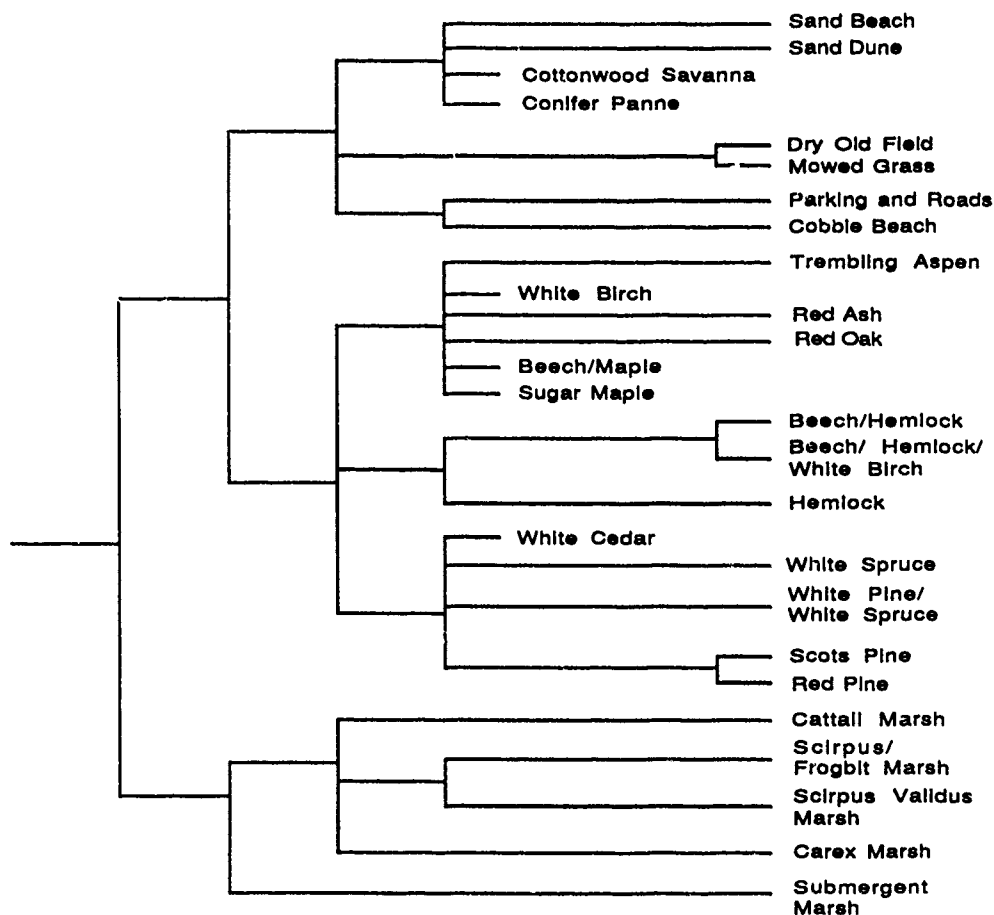


Figure 3. Classification scheme based on TWINSpan Analysis.

Accuracy Assessment

The mapping accuracy was examined in two stages. First, the classification accuracy was determined. This assessment describes the agreement between the actual class membership for the pixels in the training area and the predicted class membership derived through the maximum-likelihood classification. Second, the interpretation accuracy was determined by comparing actual class membership of the test area pixels (verification plots) to the predicted class membership arrived at by the maximum-likelihood classifier.

Classification accuracy is usually presented as contingency tables or classification error matrices. In this paper, however, space permits only a summary of individual class accuracies for one of the classifications.

Results

The subjective classification used for the 27 classes was shown in Figure 3. From the contingency table derived from the accuracy assessment, a summary has been compiled (Table 1). As can be seen, for the training area data the overall classification

accuracy is 71%, with accuracies for individual classes ranging from greater 95% to 24%. For the test area data, it was only possible to identify ground plots for 13 of the classes. Overall classification accuracy is only 40%.

Discussion

The training area data suggest that the more homogeneous and the more spectrally unique the classes are, the greater the classification accuracy is likely to be. For example, there is good differentiation of the wetland classes and also cobble beach, sand dunes, parking areas and roads. Although not shown in Table 1, there are relatively high errors of omission and commission (averaging about 40 - 50%) in both the deciduous classes and the coniferous classes, a similar result to Craighead et al. (1988).

Test area data are less conclusive. To a large extent this is because it was not possible to identify a full range of verification plots and in other cases the number of plots to perform the tests was relatively small. For example, in the training area data the conifer panne class had an accuracy of 90%, but in the test area the figure is 0%. This is because the training

Table 1 Classification Results Ranked by Order of Accuracy of Training Area Results

| Class Name | Class Accuracy (%) | |
|---------------------------|--------------------|-----------|
| | Training Area | Test Area |
| Cottonwood Savanna | 95 | 44 |
| Sand Beach | 95 | - |
| Scirpus Validus Marsh | 95 | - |
| Submergent Marsh | 92 | 14 |
| Conifer Panne | 90 | 0 |
| Scirpus/Frogbit Marsh | 89 | - |
| White Spruce/White Pine | 84 | 78 |
| Dry Old Field | 82 | 28 |
| Carex Marsh | 78 | 65 |
| Red Pine | 78 | 82 |
| Mowed Grass | 76 | 30 |
| Cobble Beach | 74 | - |
| Beech/Hemlock | 73 | - |
| Sand Dune | 73 | - |
| Trembling Aspen | 73 | - |
| Cattail Marsh | 73 | 4 |
| Parking and Roads | 71 | - |
| Red Oak | 70 | - |
| Red Ash | 63 | 43 |
| Sugar Maple | 63 | 5 |
| White Birch | 59 | 35 |
| Hemlock | 55 | 24 |
| White Cedar | 52 | - |
| Scots Pine | 51 | - |
| White Spruce | 48 | - |
| Beech/Hemlock/White Birch | 37 | - |
| Beech/Maple | 24 | - |

area data did not effectively sample the full range of spectral variability of this class and only two test plots were available to test the accuracy of the classification.

From the results of the study to date, it is observed that the best accuracies are obtained from classes that have the least variation in their surface covers, are the most homogeneous in terms of their spectral appearance at a 5 m x 5 m spatial resolution and are also ecologically unique. Thus, conifers in plantations are identified with a relatively high degree of accuracy, as are some of the wetland classes. Difficulty is encountered in mixed forest areas for two interrelated reasons. First, there is a continuous variation in species composition, stand density and understorey composition, and in turn these variations cause differences in the spectral characteristics recorded on the imagery.

Conclusions

From the results of the study, it is concluded that:

- Classes identified with the highest degree of accuracy are those where the ecological class contains a relatively small number of species and where the class has a distinctive spectral appearance.
- The less spectral variance that there is within a class, the more accurately will that class be identified.
- Emphasis needs to be placed on developing an eco-spectral classification, one in which spectral

reflectance of plants can be directly matched to plant communities.

Acknowledgements

This research was supported by an Ontario Ministry of Natural Resources grant to Drs. Suffling and Howarth and also NSERC Operating Grant A0766 awarded to Dr. Howarth.

References

1. Craighead, J.J., F.L. Craighead, D.J. Craighead and R.L. Redmond, "Mapping Arctic vegetation in northwest Alaska using Landsat MSS imagery", *Nat. Geog. Res.*, Vol. 4, No. 4, pp. 496-527, 1988.
2. Hill, M.O., "TWINSPAN, a FORTRAN Program for Arranging Multivariate Data in an Ordered Two-Way Table by Classification of the Individuals and Attributes", *Ecology and Systematics*, Cornell University, Ithaca, 90 pp., 1979.
3. Hopkins, P.F., A.L. Maclean and T.M. Lillesand, "Assessment of Thematic Mapper imagery for forestry applications under lake state conditions", *Photogramm. Eng. and Remote Sensing*, Vol. 54, No. 1, pp. 61-68, 1988.
4. McColl W.D., R.A. Neville and S.M. Till, "Multi-detector Electro-optical Imaging Scanner MEIS II", *Proc. 8th Can. Symp. on Remote Sensing*, Montreal, pp. 71-77, 1983.
5. Talbot, S.S. and C.J. Markon, "Intermediate-scale vegetation mapping of Innoko National Wildlife Refuge, Alaska using Landsat MSS digital data", *Photogramm. Eng. and Remote Sensing*, Vol. 54, No. 3, pp. 377-383, 1988.

EVALUATION OF CLASSIFICATION RESULTS OBTAINED WITH COMBINED MULTITEMPORAL OPTICAL AND MICROWAVE DATA

Alessandra Fiumara Nazzareno Pierdicca
TELESPAZIO S.p.A
Via A. Bergamini 50 - 00159 Rome - Italy

ABSTRACT

The integration of information collected by different sensors is promising for discrimination between vegetation classes. In this study two data sets with very different characteristics are digitally merged and a single data set containing both information types is generated. The two data sets are airborne SAR and LANDSAT TM data acquired over an agricultural region of Italy. Both "per pixel" and "per field" classifications are performed; results provided by LANDSAT TM alone, radar alone and by the combined data set are evaluated and compared.

1. INTRODUCTION

Past and present studies reveal the agricultural classification capabilities of optical sensors, yet these are limited by the need to have cloud-free coverage of the area under investigation. A way to bypass this limitation is to use radar, which is weather independent. Recent studies show the utility of SAR for crop type discrimination. Moreover, due to its sensitivity to the dielectric and geometric properties of vegetation, radar could improve correct crop classification rates as compared to those obtained with optical data alone. When dealing with SAR, "per field" classification is to be considered: "per pixel" classification of SAR data shows poor results due to the statistical properties of radar images (large within field variance) (Ref. 1).

In this work, a LANDSAT TM image is used together with two airborne X-band SAR images to classify a typical Italian agricultural region, where extensive ground truth is available. The first step of this work is to digitally merge information collected by the two sensors having different characteristics and generate a data set containing information from both. The second step is to evaluate "per pixel" and "per field" classification accuracies obtained with microwave data alone, optical data alone and with the combination of both.

2. DATA AND TECHNIQUES

2.1. TEST AREA AND AVAILABLE DATA

The test site is located South of the Po river, in Northern Italy. It is mainly a flat agricultural area showing a typical Italian complex parcel structure: many of the agricultural fields are less than 1 ha; the most important crops are sugarbeet, corn, wheat and alfalfa.

The radar data used in this experiment are the backscattering coefficients measured during the multitemporal airborne AGRISAR'86 campaign, by the French SAR VARAN-S (X-band, HH and VV polarization), over an area of 100 Km² of the test site, during the crop growing season (Ref. 2). The processed products are 1 look, 8 bits amplitude, 2,5 m pixel spacing images. Two acquisitions are selected for this study: June 26th, HH polarization and July 17th, VV polarization. In a previous analysis (Ref. 3) over the complete data set (four acquisitions with both polarizations), these channels were found to give the highest discriminability between agricultural classes.

For the optical portion of this study a LANDSAT Thematic Mapper scene is used, acquired over the test site, under cloud-free conditions, on July 15th. The spectral bands selected in this experiment are TM channel 3 (0.624 - 0.693 μ m), 4 (0.776 - 0.905 μ m) and 5 (1.568 - 1.784 μ m) which are sensitive to vegetation parameters (Ref. 4).

The available data set also includes ground truth data necessary for the identification of the training fields as well as for assessing the results: digital maps with fields boundaries and crop types labels are available for forty square areas of .025 Km². The sample areas were selected by systematic sampling of the site and ground truth collected in closest temporal proximity of each airborne flight.

There are no significant changes in the observed agricultural crops between end of June and mid-July. At mid-July sugarbeet and corn are still growing while wheat has been harvested but stubble is left drying on the ground; alfalfa has a short phenological cycle therefore it shows different growing stages at the same date providing high variance responses to microwave and optical observations.

2.2. DATA PREPARATION FOR ANALYSIS

The objective of this experiment requires different data processing steps before final analysis is performed.

As far as SAR images are concerned, an adaptive FIR (Finite Impulse Response) Gaussian filter is applied in order to reduce within field variations and preserve fields boundaries. This involves degradation of spatial resolution but "per pixel" classification is expected to give better results. Images are also stretched to enhance contrast and make visual interpretation easier.

An important step in combining data from different sensors is the geometric registration of the images. In this experiment the two sets of data, with the extreme difference in spatial resolution (2.5m for SAR images, 30m for TM), are formatted to the same pixel size, that is 7.5 m. SAR images are subsampled while the LANDSAT TM scene, previously registered to a topographic map, is digitally expanded by a factor of four by nearest neighbour resampling. It is not worth while expanding more: many fields of small size do not appear on the TM scene. The pixel duplication produces a blocky image; nevertheless this method is chosen because it does not alter the radiometric content of the original data. Once the images are of the same pixel size, ground control points are selected and SAR images are registered to the TM data. Nevertheless some geometric errors still remain and affect the "per pixel" classification accuracy as evaluated in this experiment and described in the following paragraph.

The different information are then digitally merged by "band replacement": a five band image is created with TM band 3, 4 and 5 and the two radar acquisitions. To reduce the computing time and the amount of data to be manipulated, the sample areas are extracted from the entire test area and a mosaic image created (28 sample areas of 67*67 pixels are recognized, the other ones lying on image borders). The digitized ground truth maps of the sample areas are combined as a further band of the mosaic image (see Figure 1).

2.3. DATA ANALYSIS TECHNIQUES

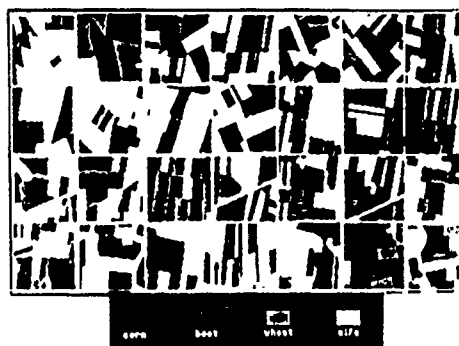
Both supervised and unsupervised "per pixel" classification are applied to the mosaic data set and the performance assessed. Several combinations of SAR and TM classification outputs are also attempted. Classification results are usually evaluated with testing points or areas. In this experiment, correct classification rates are computed on a "correct classified" image obtained as the logical "and" between the output of the classification algorithm and the ground truth map.

Substantial improvement in classification accuracy is expected, in particular for the radar, if field averages rather than pixel values are used in the classification. The improvement is expected to be smaller for LANDSAT. Since the development of image segmentation techniques is not the aim of this study, the ground truth map superimposed on the five band image is used instead of the result of a segmentation algorithm or an available large scale map to select homogeneous fields. a measure of class discriminability, the divergence index is computed with the field mean values. The divergence is a measure of how separable two classes are, given their covariance matrices and mean vectors (Ref. 5). It is computed for each class pair and for every combination of bands to identify the classes with poor separability as well as the channels useful for discriminating between them. A classification procedure is also carried out over the same set of fields mean values: for every channel combination, a discriminant function is developed for the data set using the generalised squared distance (Ref. 6) and the pooled covariance matrix; the classification criterion is applied to the data and the confusion matrices generated and compared using the K coefficient (Ref. 7). The estimate of this coefficient and its variance are used to evaluate and compare (within a given confidence interval) the classification results. Since in this experiment the same data are used for training and testing, the results can not be considered as an absolute accuracy index but as a measure of relative variations.

Figure 1: The mosaic data set
a: SAR July 17th, VV polarization



b: Digitised ground truth map



3. RESULTS AND DISCUSSION

3.1. "PER PIXEL" CLASSIFICATION

Training data statistics are extracted for 31 fields: 7 fields of corn, 9 of sugarbeet, 5 of alfalfa and 10 of wheat. Then, the maximum likelihood classifier is run over a total of 97559 pixels. The correct classification rate obtained with SAR multitemporal data set alone is poor (54.37%). The classified map shows high confusion in particular between pixels of alfalfa and corn, while better results are obtained for sugarbeet and wheat.

The addition of optical information improves correct classification rate, as expected. However, classification techniques applied to the five band image are always affected by the large within field variance of SAR data and produce comparable accuracies.

Different combinations of classification outputs of SAR and LANDSAT are tested separately: the best classification map is obtained when SAR is used to classify sugarbeet and LANDSAT used for the other crops. The resulting correct classification rate is 67.67% which is however comparable to the use of TM alone.

3.2. "PER FIELD" CLASSIFICATION

Fields averages are extracted for 161 fields: 24 fields of corn, 39 fields of sugarbeet, 58 fields of wheat and 40 fields of alfalfa. Boundaries recognition is done by means of the digitised ground truth map.

The results of the divergence computation for each class pair are listed in Table 1. Class divergence is ranked as a function of the number of channels used. The channel set with the highest separability index is selected for each class pair.

The critical class pair for SAR is wheat-alfalfa: even with multi-date SAR channels, the divergence index is very low (4.97). Optical data increase the separability between these two classes: stubble fields and vegetated areas appear very different on a LANDSAT image. The divergence index becomes 47.67 when only one TM band is used (TM band 5).

Optical data show poor capability in discriminating alfalfa from corn and from sugarbeet. SAR is revealed being useful for the separability of such classes: an improvement of 93.33% of the divergence index between alfalfa and corn is obtained when SAR June 26th data are added to TM band 5. For the class pair alfalfa/sugarbeet, the use of SAR image June 26th improves the divergence index from 1.54, obtained with TM band 3 or 9.27 obtained with TM band 4, to 23.80. The results obtained with radar data alone are still higher (42.41) than the results obtained with TM even when two optical channels are used (the maximum value is 18.39 with TM4 and TM5).

Significant increments in overall interclass divergence (see ΔDiv in Table 1) are generally observed with two multi-sensor channels. This

is also confirmed by the K coefficient estimates. KHAT coefficient and confidence interval are presented in Figure 2. The highest value of KHAT is given by LANDSAT TM band 5 and SAR June 26th, HH polarization for the two channel case. If a 0.05 confidence interval is assumed, the results of TM5+S1 are significantly higher than those of TM4+TM3 and TM5+TM3. When three channels are considered, multisensor sets have to be preferred; in particular TM4+TM5+S1 shows the best KHAT value, although the difference from the others is not significant.

4. CONCLUSIONS

The major conclusions of this study are the following. No significant improvement in classification accuracy is obtained by adding optical information to X-SAR data if the classification is performed "per pixel". On a field basis, the results reveal that the two sensors are complementary and that both information have to be combined to discriminate between particular classes. No more than two channels (one from SAR and the other from Thematic Mapper) are needed to achieve acceptable classification accuracy. The results obtained in this study are related to cover types, geographic location and time period specified.

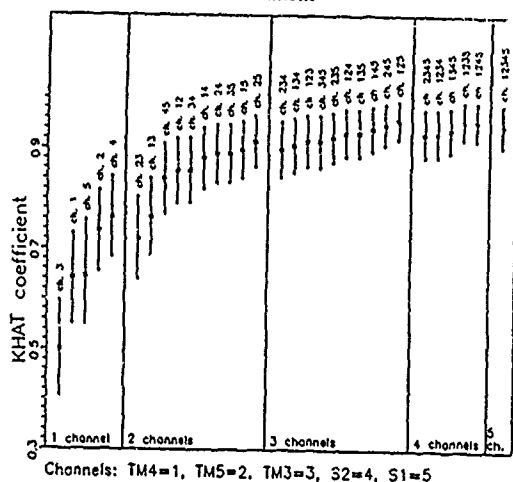
ACKNOWLEDGEMENT

This work has been supported by the Italian Space Agency in the framework of the Remote Sensing Pilot Project.

Table 1:
Class separability as a function of the number of selected channels.
Crops: b=sugarbeet, c=corn, a=alfalfa, w=wheat
Bands: S1=SAR June 26th/HH, S2=SAR July 17th/VV
TM3=TM band 3, TM4=TM band 4, TM5=TM band 5

| class pair | N° ch | Max.Div | ΔDiv | selected channels |
|------------|-------|---------|--------------|-----------------------|
| c/b | 1 | 27.03 | - | TM4 |
| | 2 | 37.16 | 37.47 | TM4,S2 |
| | 3 | 45.28 | 8.12 | TM4,S1,S2 |
| | 4 | 49.57 | 0.09 | TM4, TM5, S1, S2 |
| | 5 | 55.92 | 0.12 | TM3, TM4, TM5, S1, S2 |
| c/w | 1 | 44.38 | - | TM5 |
| | 2 | 68.23 | 23.85 | TM4, TM5 |
| | 3 | 93.55 | 37.10 | TM4, TM5, S2 |
| | 4 | 101.73 | 8.74 | TM4, TM5, S1, S2 |
| | 5 | 108.3 | 6.45 | TM3, TM4, TM5, S1, S2 |
| c/a | 1 | 8.55 | - | TM5 |
| | 2 | 16.53 | 93.33 | TM5, S1 |
| | 3 | 20.92 | 26.55 | TM5, S1, S2 |
| | 4 | 24.4 | 16.63 | TM4, TM5, S1, S2 |
| | 5 | 25.30 | 3.68 | TM3, TM4, TM5, S1, S2 |
| b/w | 1 | 108.21 | - | TM3 |
| | 2 | 193.99 | 79.27 | TM5, S2 |
| | 3 | 255.15 | 31.52 | TM4, TM5, S1 |
| | 4 | 287.60 | 12.71 | TM4, TM5, S1, S2 |
| | 5 | 318.93 | 10.89 | TM3, TM4, TM5, S1, S2 |
| b/a | 1 | 23.80 | - | S1 |
| | 2 | 42.41 | 78.19 | S1, S2 |
| | 3 | 63.28 | 49.21 | TM5, S1, S2 |
| | 4 | 86.47 | 36.64 | TM4, TM5, TM3, S1 |
| | 5 | 92.47 | 6.93 | TM3, TM4, TM5, S1, S2 |
| w/a | 1 | 47.67 | - | TM5 |
| | 2 | 68.31 | 43.29 | TM3, TM4 |
| | 3 | 95.42 | 37.66 | TM3, TM4, TM5 |
| | 4 | 102.29 | 7.19 | TM4, TM5, TM3, S2 |
| | 5 | 107.80 | 5.38 | TM3, TM4, TM5, S1, S2 |

Figure 2: Estimated K coefficient as a function of channel combinations



REFERENCES

1. Ulaby f., R. Li, K. Sanmugan, 1982, "Crop Classification Using Airborne Radar and Landsat Data", IEEE Trans. Geoscience and Remote Sensing, Vol.Ge.-20,No.1.
2. AGRISAR'86 Preliminary Investigators' Report, 1987, ed. A.J. Sieber
3. Fiumara A., N. Pierdicca, M. Ricottilli, 1988, "Crops Radar Responses Analysis Based on Agrisar'86 Data", Proceedings of IGARSS'88.
4. "Applications of Remote Sensing to Agrometeorology", Proceedings of an ISPRA course on Remote Sensing, 1987, ed. F.Toselli.
5. Thomas I., V.Benning, N.Ching, 1987, "Classification of Remotely Sensed Images", Adam Hilger, Bristol.
6. Rao, C. Radhakrishna, 1973, "Linear Statistical Inference and its Applications", New York: John Wiley & Sons
7. Congalton R.G. Oderwald and R.A. Mead, 1983, "Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques, PE&RS, vol.49, N.12, pp.1671-1678

COVER CLASS MAPPING IN AGRICULTURAL ENVIRONMENTS

N A Campbell and J F Wallace

Division of Mathematics and Statistics, CSIRO
Wembley 6014 W.A. AUSTRALIA

Tel: (619) 387 0294 Fax: (619) 387 6046

ABSTRACT

The paper outlines some collaborative remote sensing studies undertaken by the CSIRO Division of Mathematics and Statistics. The studies have been concerned with the mapping of pixels into various land cover classes, in both agricultural and rangeland settings. The problems of assigning pixels to one or other of a number of reference classes, such as crop, pasture and bush in agricultural areas, and/or determining the proportions of cover types within a pixel, such as bare soil and shrub in rangelands, are common to most of the projects.

The agricultural studies have focussed on mapping the areas sown to crop and pasture; delineating areas affected by primary and secondary salinity; identifying vegetation classes within nature reserves; and detecting areas of crop affected by waterlogging.

The statistical and collaborative projects carried out by the group have resulted in the development of a simple-to-use image processing system based on the Commodore Amiga; the package is now available commercially.

Keywords: cover class mapping; canonical variate analysis; classification.

BACKGROUND

In 1984, the CSIRO Division of Mathematics and Statistics (DMS) established a research project on the Analysis of Remotely Sensed Data. This

was in response to DMS collaboration in several agriculturally-based remote sensing studies, and the perceived need to carry out research into statistical problems arising from this collaboration.

The general objectives of the Remote Sensing Project have been to assess, develop and implement methods for the analysis of remotely sensed data and to collaborate in research projects with other CSIRO Divisions and appropriate Government and industry groups.

The collaborative studies have generally been concerned with the mapping of pixels into various land cover classes, in both agricultural and rangeland settings. Assigning pixels to one or other of a number of reference classes, such as crop, pasture and bush in agricultural problems, and/or determining the proportions of cover types within a pixel, such as bare soil and shrub in rangeland problems, are common to most studies.

The major agricultural studies have focussed on mapping the areas sown to crop and pasture, delineating areas affected by primary and secondary salinity, identifying vegetation classes within nature reserves, and more recently, detecting areas of crop affected by waterlogging. The various studies are interrelated; distinguishing areas sown to crop from those in pasture is necessary before the effect of waterlogging can be evaluated.

CROP FORECASTING

In 1981, DMS undertook a collaborative project with Cooperative Bulk Handling (CBH), the local grain handling authority, to evaluate the role of Landsat multispectral scanner (MSS) data in crop forecasting in the wheatbelt of Western Australia (WA). Interest was to be focussed on the evaluation of areas sown to crop and on discrimination between crops.

The results discussed here are based on analyses of multi-temporal Landsat MSS data for the Corrigin district for the 1981 growing season (path 119; row 82). Seven cloud-free images were available for this scene between June and early November.

Initial analyses of data from several farms showed good spectral separation between crop and pasture. For example, analyses of data from selected paddocks from one property showed better than 95% accuracy of identification of wheat and pasture pixels using three overpasses in July and August. However, three-colour displays of selected combinations of bands (e.g. of band 7 - band 5) for these overpasses, and related statistical analyses, indicated that the criteria established on one farm could not be used with any degree of confidence to identify pixels as crop or pasture on other farms.

The traditional approach, in Australia and elsewhere, has been to use selected paddocks within regions of the order of 5-10 km square to establish suitable statistical criteria for crop identification and then carry out a classification for the region. Given the difficulty of extrapolating results from farm to farm, such an approach would not be successful in the WA wheatbelt.

How, then, could Landsat data be used for crop forecasting? In view of the promising separation between crop and pasture for the initial farms and the ready identification of individual farms on Landsat MSS imagery, it was decided to see if the farm itself could provide the basic sampling unit.

Farm plans giving details of areas sown to crop and pasture in 1981 were obtained for twenty-four test farms in the Corrigin district. Some farm plans were extremely comprehensive, giving details of variety of crop or pasture, date of sowing and soil type, while others simply indicated whether a paddock was in crop or pasture. To enable the data from the overpasses to be analyzed together, the satellite data were registered to the overpass on August 30, using a quadratic polynomial and nearest-neighbour resampling.

Training areas consisting of between twenty and one-hundred or so pixels which were representative of a single cover type (e.g. spray-seeded wheat, conventionally-seeded wheat, ungrazed pasture, heavily-grazed pasture) were selected.

Canonical variate analyses of the spectral signatures for the Landsat MSS data for the various farms showed a broadly similar pattern - training areas representing bush showed a distinct signature, as did training areas selected from bare/saline areas. Since the dark bush areas and the bright bare-to-saline areas were readily distinguishable both visually and statistically, the analyses concentrated primarily on the remaining areas of the farm.

Canonical variate analysis makes no assumptions as to whether a training area within a paddock represents crop or pasture; the analysis simply treats each training area as a distinct group and seeks to establish the degree of spectral similarity or difference between the groups. The paddock labels provided by the farmer are added to the resulting plots, to see if the groupings established by the statistical analysis correspond to cover classes representing crop and pasture.

In general there was consistent separation between paddocks representing wheat and paddocks representing pasture within a farm. The results from this more detailed evaluation confirmed the observation that the results from one farm could not necessarily be used to label pixels on other

farms. For approximately half the study farms, the overall spectral separation between crop sites and pasture sites could be summarized by a simple spectral index, while for the remaining farms the nature of the spectral separation changed from date to date.

In conclusion, while good spectral separation could be established between crop sites and pasture sites within nearly all of the farms in the study, the variation in the nature of the separation from farm to farm makes it necessary to extract training sites and carry out the analyses separately for each farm.

MAPPING DRYLAND SALINITY

The WA Department of Agriculture and CSIRO have been collaborating on a project to examine the feasibility of using Landsat MSS data to identify and map saline areas in the wheatbelt of Western Australia.

The analyses reported here are based on three satellite overpasses for the Moora region (8 July, 31 August and 23 September, 1981) (path 120; row 81). The data were rectified to AMG coordinates using a quadratic polynomial relationship and nearest-neighbour resampling to a pixel size of 56 m x 79 m.

Initially, one hundred training sites covering the range of spectral variation evident in three-colour displays were extracted. A plot of the site means for the first few canonical variates was then used to define segments containing sites with similar spectral variation. One or two sites within each segment were selected as reference classes in a classification procedure, and the remaining sites were allocated to one or another of these reference classes. Those training sites for which most of the pixels were atypical were studied more closely on aerial photographs, three-colour displays and the canonical variate plots and further reference classes were identified. Several iterations of this approach resulted in twenty-three reference classes being defined.

The classification performance was then examined in the context of regional mapping. Four test sites were selected, each one consisting of regions of bush, crop and/or pasture, and saline areas. Corresponding ground information was obtained from photo-interpretation of enlarged 1:50000 aerial photographs. The resulting boundaries were captured using the Intergraph Interactive Graphics and Digitising System. Polygons were formed using the captured boundaries, each polygon having a linkage to a database containing attributes for the photo-interpretation classification code, area and perimeter. A grid with a cell size of 59 m x 79 m and an origin corresponding to the resampled Landsat MSS images was generated for each of the four test sites. The class code for the predominant class was then assigned to each cell. The computer classification was then compared with these ground maps.

Overall, the four test sites contained 2416 pixels mapped as salt from the photo-interpretation. Of these, 1956 pixels (80%) were allocated as salt, while 460 pixels (19%) were not. A further 224 pixels (equal to 10% of those mapped) were classified as salt but were not labelled as salt.

The spectral allocation of boundary pixels containing a mixture of classes created some confusion. The production of the ground maps also suffered from the same difficulty, since the boundary position is critical in relation to the pixel grid. However, disagreements of this kind showed up as single pixels along boundaries, and so could be accommodated in any summary.

The major concern was with blocks of pixels which were wrongly allocated. Usually, such pixels were also atypical (relative to the current reference classes), indicating the need for additional reference classes. Inconsistencies between the labelling of the ground maps for the test sites and the labelling of the original training sites also caused some difficulties since the selection and tentative labelling of the training sites, and the photo-interpretation of the test sites, were carried out by different people.

It is important to note that these results were obtained from a first iteration of the classification procedure for the test sites. Examination of the classification maps has indicated the need for further definition and refinement of the reference classes before a large-scale regional mapping is undertaken.

IDENTIFICATION OF AREAS OF WATERLOGGING

The Remote Sensing Project (in collaboration with the WA Department of Agriculture) has been investigating the degree to which areas of waterlogging can be identified from multispectral airborne scanner data.

An initial evaluation was carried out using data obtained from a flight of the GEOSCAN scanner on 6 September 1986, the flight path being chosen to include several test farms from the CSIRO-CBH Crop Forecasting Project. (The scanner has three spectral bands in the visible region, two in the near-infrared, four in the mid-infrared, and three thermal bands.) One of the farms was visited subsequently by a field officer from the Department of Agriculture and areas of waterlogging were recorded.

Analyses of training sites from various parts of the farm showed clear separation between sites extracted from the main waterlogged area and those representing healthy crop and pasture. An evaluation of the contribution of the various spectral bands showed that a near-infrared band (830-870 nm), two mid-infrared bands (1980-2080 nm, and 2300-2400 nm), and two thermal bands (8.50-9.00 μm , and 9.70-10.20 μm) effected the separation. Waterlogged sites showed higher responses in the 2350 nm band relative to the 850 nm and 2030 nm bands, and in the 9.95 μm band relative to the 8.75 μm band.

A classification based on these five spectral bands identified a region of waterlogging consistent with that recorded on the ground maps. The classification procedure was also applied to three smaller areas visited by the field officer, and again indicated waterlogging in regions identified on the ground.

As a result of this initial investigation, a more detailed study was undertaken in 1987. This latter study focussed on the Upper Great Southern Region of WA. The test sites were flown on July 25, while colour aerial photographs were obtained in early May and in mid-September.

The findings from the 1987 study were similar to those for the 1986 evaluation, namely that airborne scanner data could be used to map waterlogged crops, and that the spectral separation was effected by the same near-infrared, mid-infrared and thermal bands.

VEGETATION CLASSIFICATION IN NATURE RESERVES

Landsat MSS data have been used to provide information on broad vegetation classes within nature reserves in the wheatbelt of Western Australia.

The analysis was carried out using a summer (December 1980) and winter (August 1981) Landsat MSS overpass, the December pass being registered to the August pass using a third-degree polynomial with 20 ground control points. A number of training sites (~43) were then selected, each one chosen to lie wholly within mapped vegetation units.

An overall brightness index separated the lighter vegetation from the heath and shrubland and woodland and mallee sites, while a vegetation index contrasting the visible and infra-red bands separated the woodland and mallee sites from heath and shrubland. Examination of the important subsets of bands showed that an analysis based on a single date gave substantially poorer separation than a two-date analysis, and that discrimination based on Landsat bands 4, 5 and 7 weighted for each overpass was adequate.

Allocation of pixels into classes representing sparse and open heath, Casuarina heath and shrubland, and mallee and woodland, compared well with the results from ground and photographic surveys.

THE USE OF MULTIDATE LANDSAT THEMATIC MAPPER IMAGERY
FOR AGRICULTURAL LAND USE MAPPING
APPLIED TO ELECTRIC POWER TRANSMISSION ROUTE SELECTION STUDIES

R.N. Pierce and C.L. Wagner

Land Use and Environmental Planning Department
Ontario Hydro
Toronto, Ontario

ABSTRACT

Agricultural landcover data was derived from Landsat TM imagery using computer assisted image analysis techniques. This data contributed to an environmental assessment for establishing route locations for high voltage transmission facilities in southern Ontario. Using images from two stages of the growing season it was possible to reliably interpret the individual cover types required to compile a digital agricultural land use systems map. This method provided an efficient and cost effective alternative to manual interpretation methods and extensive field work. These results have established the use of this technology for future route and site selection projects.

Key Words: Agriculture, Landcover, Landsat TM, Multidate, Ontario Hydro

INTRODUCTION

The mapping of agricultural crop cover data is an integral part of the environmental impact assessment process for locating electric power transmission facilities at Ontario Hydro. Historically, this information has been gathered using conventional aerial photography, Ministry of Agriculture inventories and field observations. Recently, computer assisted analysis of satellite imagery has been able to provide some of this data. Agricultural cover types mapped in digital form can be used directly by planners involved in preparing impact assessment models.

The focus of this project was to interpret crop types and other vegetative cover for a study area defined by a proposed bulk power transmission route location west from London Ontario. The data derived from the image analysis would be integrated with other land use information to identify agricultural land use systems for an environmental impact constraint map. The work was carried out in the spring of 1988 for use in identifying alternative route locations scheduled for the summer of 1988. From previous studies (Pierce, 1987) and the literature (Hixson et al, 1978; Jackson, 1986) it was expected that the method proposed would be suitable and cost effective.

METHOD

The approach taken for this project was based on preliminary work carried out using established computer assisted image analysis procedures. Discussion with staff involved in the route selection project determined the set of landcover types required. Images were acquired, processed and the results were checked with the ground truth data. The final product was then transferred to an in-house Geographic Information System (GIS) and incorporated in the planning process.

Study Area

The area under study extends from London in the northeast to a western boundary formed by Sarnia, the St. Clair river, Lake St. Clair, and Windsor. The southern boundary runs from the Detroit River along the shore of Lake Erie encompassing a total area of approximately 11,000 sq. kilometres. The physiography of the area is varied with major units of clay plains, sand plains and till plains. Land use is dominated by agriculture with a full range of intensities, field patterns and crops. Forest vegetation is primarily deciduous, located in small scattered woodlots and river valleys.

Data Sources

Landsat 5 Thematic Mapper (TM) data was the best available source of digital imagery. The characteristics of TM data provide sufficient resolution to separate individual fields which can be identified on the ground control sources. Temporal resolution was a factor in the selection of appropriate imagery. It was necessary to identify the period in the growing season that would best represent the reflectance values for the crops of interest. It was determined that a late August date would provide the best coverage of corn and soybean crops. Specialty crops such as tomatoes and other canning crops would require a midsummer date and winter wheat and hay/pasture would require an early season (spring) image.

Based on an image search with the criteria of minimal cloud cover and the need for up-to-date information, it was possible to locate spring and late summer images. Unfortunately, a suitable midsummer scene could not be obtained and the ability to identify specialty crops was therefore

limited. Seven band, bulk, full scenes were acquired for track 19, frame 30+10, for April 23, 1986 and August 29, 1986.

Detailed ground truth was required for the image analysis procedure. Individual field crop audit information for 1986 was provided by the Ontario Ministry of Agriculture and Food (OMAF). This information was contained in random blocks approximately 2000 ha in size for all five counties located within the study area. Aerial photography, aerial video surveys and Agricultural Land Use System maps (OMAF) were also used for ground truth.

Image Processing and Classification

All image processing was carried out using a Dipix ARIES III image analysis system. Raw image data was resampled to conform with a standard UTM geographic base. The spring image was registered to the summer image in order to merge the results of the analysis from both dates. The resampled image was approximately 3200 lines by 5200 pixels, requiring 18 megabytes of disk space per file.

Based on preliminary examination and a trial classification, the summer image appeared to have the most potential for identifying the majority of the land cover types. However, there appeared to be regional variations in crop type and pattern. These tended to correspond with natural and physiographic patterns within the study area. In order to minimize misclassification of crops within these variations, the summer image was divided into two regions based on observed changes and known physiographic features (Chapman and Putnam, 1966). Thus, two sets of analyses for the summer image were performed and the results from the two regions were combined in the final map.

Standard supervised classification procedures were carried out using the ARIES image analysis software. TM bands 4,3 and 2 were displayed (RGB) to perform the interpretation on both the spring and summer images. Training areas were created from the summer image, using the OMAF crop audit data, for corn, soybean, grains, tobacco, fallow fields and bare soil, and deciduous and coniferous forest. Signatures were generated from bands 1,2,3,4,5 of the TM data. A maximum likelihood classification was performed and the results were compared with the OMAF information. Forest cover was compared with Ministry of Natural Resources Forest Resource Inventory maps and aerial photos.

Initial observations indicated that the first classification produced a reasonable representation of the required cover types. Problems encountered in the summer scene included unclassified areas, crop inclusions and some misclassification. The training areas were refined so that more representative values would be included in the signatures. Several signatures were created for grain due to the large variation in the appearance of the fields. Each signature or combination of signatures was tested for the amount of grain field included in the classification. It was found that two signatures for this cover type would be adequate. A signature was also added for areas of urban development. Subsequent iterations of the classification were performed varying the statistical rejection level for several of the classes.

The results from the above modifications showed improvement in the representation of the cover types. However, due to the high standard deviation values for the urban development category, this class overlapped with a number of the other cover types. Consequently, an attempt was made to extract this information from only one spectral band. Band 3 of the summer scene provided the best spectral separability and was therefore used in a parallelepiped procedure to produce the urban development theme.

Training sets for hay, pasture and winter grains were developed from the spring image. These were easily identified due to the lack of any other green vegetation growing at this time of year. Autocorrelation of the three signatures derived from the spring image showed that hay, pasture and winter wheat were spectrally separable using this date of imagery. A maximum likelihood classification (also based on bands 1 to 5) was performed on the image.

Final versions of the classifications of both the spring and summer images were selected. The two regions of the summer image were merged and the spring and summer themes were combined using a "logical operations" task. At this time as well, a theme for water, and the urban development cover type were included to complete the landcover map. Some minor classification errors were interactively modified using a manual theme generation function. In addition, a post classification filter of 10 pixels was used to eliminate individual and small clusters of extraneous themes and unclassified pixels.

RESULTS

The product of this process included two theme files with a total of nine themes: corn, soybean, grain, grasses (hay and pasture), tobacco, deciduous forest, coniferous forest, urban development, and water. Maps were output to a colour plotter for selected areas and the cover types were checked for accuracy of classification using the original agricultural audit data as well as a second set that was not used to develop training areas. It was observed that the results of the classification corresponded well with the agricultural audit data.

Field investigation conducted in early April 1988, confirmed classification results for areas of permanent pasture, grasses and idle land; crops that are not on a rotational cycle; and physical features such as drainage and topography that affected the signatures. It was also possible to observe the stage of growth of the winter grains and grasses relative to the reflectance characteristics observed on the April image.

Unbiased ground control information (audit data not used for training areas) was used to prepare a simple statistical assessment of the classification accuracy (see Table 1). The results indicate that most of the individual classes have an accuracy greater than 90%. The overall performance of the classification is 95.5%. Anomalies in the classification relate to confusion or misclassification with grains, grasses and urban development. This was attributed to similarities in the spectral characteristics of the features used to define the signatures. That is, in almost all areas of confusion, "dry" grass type characteristics were

present, either in the form of recently cut hay and grain and fallow fields, or in open fields and grass areas adjacent to urban areas, idle land, road rights-of-way. It may have been possible to compensate by increasing the range of grass cover types in the spring image. However, it is expected that the date of the spring image may have been too early in the growing season to record the optimum infrared reflectance of the pasture lands.

Anomalies identified in some of the other cover types can be attributed to the growth of volunteer crops in rotational fields, and local variations in topography resulting in bare patches of soil due to erosion or poor drainage. These situations were considered to be minor in light of the overall accuracy and the final application of the data.

APPLICATION

The product of this project was a map of specific crop cover types for a large geographic area of southwestern Ontario. The purpose for preparing this data was to apply this data to an impact prediction model used to assess the potential impact of high voltage transmission structures on agricultural land. Since this data was derived from imagery acquired two years prior to the application date it was apparent that there would be differences in the actual field by field crop cover due to the practice of crop rotation. However, assuming that there has been no overall change in farming practices, the fields will continue to be in a constant state of rotation. To compensate for this, a crop "systems" approach was taken whereby patterns are identified according to predominant crop types.

In this application the crop cover information was transferred to the GIS and combined with property boundary data. A model was developed that would identify specific crops in each farm unit and categorize that unit according to an "Agricultural Land Use System." The results were then used in conjunction with other agricultural data to develop an overall agricultural constraint map.

The forest cover data was used to update and verify other forest resources data that was applied to wildlife habitat impact prediction.

CONCLUSION

The method as described proved to be a suitable means for deriving agricultural landcover data. The use of multitemporal imagery demonstrated significant value for specific cover type separation. As well, the availability of historical ground cover data was a key to the success of the analysis. Based on the application of the results of this project it is concluded that considerable time and cost has been saved over the use of conventional methods. The integration of digital satellite derived data with other digital data in a GIS therefore has been established as a viable and cost effective method for aspects of high voltage electrical transmission route and site selection.

REFERENCES

- Chapman, L.J., D.F. Putnam, 1966. "The Physiography of Southern Ontario". The Ontario Research Foundation, University of Toronto Press. Map No.2225.
- Hixson, M.M., M.E. Bauer and L.L. Biehl, 1978. "Crop spectra from LACIE field measurements"; LACIE-00469, JSC-13734, and LARS Contract Report 011578. In Manual of Remote Sensing, Second Ed., R.N. Colwell, Editor in Chief, ASPRS, 11983.
- Jackson, G., 1986. "Remote Sensing and the Agricultural Resource Inventory"; proceedings, Tenth Canadian Symposium on Remote Sensing, Edmonton, Alberta.
- Pierce, R.N., 1987. "Agriculture Land Use Mapping in Southwestern Ontario Using Computer Assisted Image Analysis". Unpublished report prepared for the Land Use and Environmental Planning Department, Ontario Hydro.

TABLE 1

Confusion Matrix for Classification Accuracy

| Known Category | No. of Pixels | Percent Correct | Number of Pixels Classified into Category | | | | | | | | | |
|----------------|---------------|-----------------|---|-------|------|------|------|-----|------|------|-------|-------|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Uncl. |
| 1 Corn | 18598 | 94.6 | 17592 | 386 | 331 | 121 | 0 | 1 | 166 | 0 | 0 | 1 |
| 2 Soybean | 30672 | 95.1 | 1178 | 29159 | 7 | 319 | 0 | 0 | 6 | 0 | 0 | 3 |
| 3 Grain | 8952 | 93.5 | 151 | 224 | 8372 | 151 | 0 | 0 | 2 | 0 | 0 | 52 |
| 4 Grass | 3728 | 70.9 | 80 | 280 | 665 | 2643 | 0 | 0 | 0 | 0 | 0 | 60 |
| 5 Tobacco | 1159 | 90.3 | 5 | 2 | 31 | 0 | 1047 | 0 | 3 | 46 | 0 | 25 |
| 6 Coniferous | 417 | 98.8 | 0 | 0 | 2 | 0 | 0 | 412 | 3 | 0 | 0 | 0 |
| 7 Deciduous | 4856 | 96.0 | 30 | 104 | 5 | 3 | 0 | 15 | 4663 | 0 | 0 | 36 |
| 8 Developed | 13049 | 72.2 | 25 | 0 | 2089 | 17 | 0 | 0 | 0 | 9424 | 25 | 1469 |
| 9 Water | 97533 | 99.9 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 97518 | 10 |

THE USE OF MULTI-TEMPORAL TM TASSELED CAP FEATURES FOR LAND USE MAPPING IN EUROPEAN MARGINAL AREAS AN OPERATIONAL APPROACH

Joachim Hill Jacques Megier

Institute for Remote Sensing Applications
Lab. for Image Processing

Commission of the European Communities
Joint Research Center
I-21020 Ispra (Va), Italy

Abstract

JRC Ispra is conducting a land use inventory and mapping experiment over the Departement Ardeche (France), which is considered representative for the so-called "Less-Favored Areas" within the European Communities.

The results of a uni-temporal classification of Landsat-5 TM data have confirmed that the spectral information from single-date imagery contains a high level of spectral confusion between cover types. It was expected that better mapping accuracy can be achieved with multi-temporal approaches.

Our multi-temporal data set includes four TM scenes from the year 1984, all of which have been corrected for atmospheric absorption, scattering and pixel adjacency effects. Geometric registration was completed with sub-pixel accuracy. Time series of NDVI and Tasseled Cap features for important cover types were analyzed and it was found that the multi-temporal vegetation index alone does not provide sufficient information for the classification in complex landscape units as the Ardeche. Tasseled Cap features, however, are considered a more efficient compromise between data reduction and the preservation of relevant image information. They are especially suited for unsupervised approaches since the principle of this transformation is related to physical concepts of optical remote sensing.

Classification was performed by using simple parametric techniques such as minimum euclidean distance mapping. It is demonstrated that the multi-temporal approach leads to distinctly improved results and allows detailed ground cover mapping even in structurally complex landscapes.

Key Words: Multi-temporal TM data, atmospheric corrections, geometric registration, NDVI, Tasseled Cap features, euclidean minimum distance, land cover mapping.

1. Introduction

In 1984, a collaborative project between JRC and the statistical service of the French Ministry of Agriculture (SCEES) was initiated in order to evaluate the suitability of TM and SPOT data for agricultural inventories and natural vegetation mapping in so-called "Less-Favored Areas" of the European Communities. The study region is the Departement Ardeche in the south of France. It is a mountainous area with a large variety of natural vegetation types. The agricultural land is mainly characterized by its strong fragmentation and small parcel dimensions (0.5 - 2 ha).

In a first experiment, only one TM data acquisition from 5 July 1984 had been used to classify and map the ground cover classes of interest. An acceptable accuracy was found for the global inventory, but there were indications of bad precision in local cover mapping (Hill & Megier, 1988). Multi-temporal approaches were expected to provide better accuracy in both statistical inventory and thematic mapping, but they require reliable geometric and radiometric corrections of the satellite data.

2. Data sets and study region

The inventory experiment for the year 1984 uses four Landsat-5 TM scenes, which were acquired at 24 April (P196 R29), 5 July (P197 R29), 30 July (P196 R29) and 31 August (P196 R29). Digital elevation data are available which cover the whole administrative region with a ground resolution of 250 x 250 m.

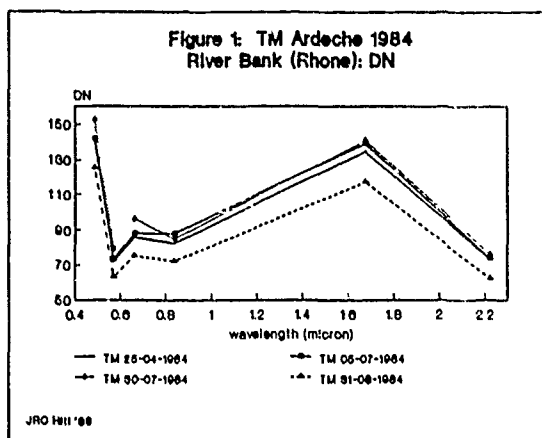
For this case study, a test area has been selected for which an almost complete coverage with aerial photographs is available. It includes a complex mixture of natural vegetation and agricultural cover types. Deciduous and coniferous forests extend over varying morphological units; low wood- and shrub lands are mainly found within the sub-mediterranean "Garrigue" and abandoned areas of agricultural terrace cultivations. The agricultural land includes cereals (mainly winter wheat), rape seed, maize and sorghum, sunflowers, alfalfa and, mainly within the Rhone valley, various orchard cultivations.

3. Geometric registration

It is evident that highly precise geometric pre-processing is required in order to fully exploit the spatial detail of multi-temporal satellite imagery. Landsat TM data are of exceptionally good geometric quality which, in level terrain, allows geodetic rectifications to sub-pixel accuracies by using only affine transformations.

The local geometry of TM scenes, however, can be considerably deteriorated by horizontal displacements of ground features which are due to terrain elevation. The effect is not negligible since terrain features in an altitude of 2,000 m above the reference ellipsoid will be displaced by about 9 pixels at each end of a TM scan. Hence it will cause misregistration of multi-temporal overlays, especially when data from different TM orbits are used as in our case.

Therefore a relief correction module was integrated into the geometric correction scheme which combines polynomial transformations and the



altitude information as made available from digital elevation data. The method guarantees sub-pixel accuracy and provides map compatible geo-coding of TM data without local distortions (Kohl & Hill, 1988). Since sufficient accuracy can be achieved by the use of commercially available elevation data of low spatial resolution (250x250 m grid) the correction method is considered operational.

4. Atmospheric corrections

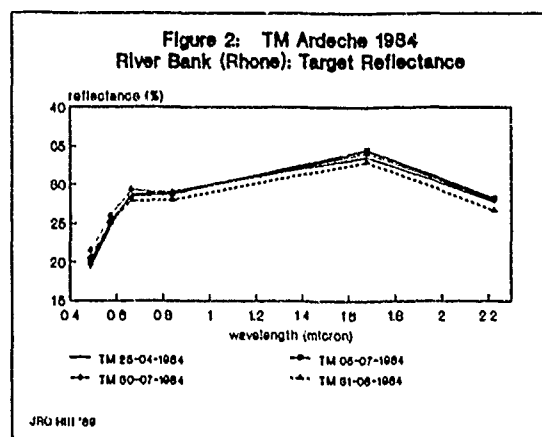
Ideally, a comparison of several scenes would only show changes in the intrinsic properties of targets and be invariant to other effects. Since the sensor measurement results from the interaction of several factors, the effects of surface material (reflectance factors) must be separated from the effects of topography, illumination, shadows, viewing directions and atmospheric path phenomena.

With respect to the low resolution of the elevation data (250 x 250 m), no correction of topographic effects was considered, but the four Thematic Mapper scenes used throughout the experiment have been corrected for atmospheric absorption, scattering and pixel adjacency effects in order to retrieve the target reflectance factors. The atmospheric correction model is based on the formulation of radiative transfer as developed from Tanre et al. (1979, 1987):

$$\rho^* = t_{03} \left\{ \rho_{at} + \frac{T(\mu_0) [t_d(\mu) \rho_t + t_s(\mu) \langle \rho \rangle]}{1 - \langle \rho \rangle s} \right\} \quad (4.1)$$

where ρ^* is the apparent at-satellite reflectance, ρ_t the target reflectance factor, $\langle \rho \rangle$ the background contribution to the apparent reflectance and s is the spherical albedo. ρ_{at} denotes the intrinsic signal component, $T(\mu_0)$ the total downward, $t_d(\mu)$ the diffuse and $t_s(\mu)$ the scattered upward transmittance; t_{03} gives the ozone transmittance. The aerosol optical thickness τ_a , which is considered the key parameter for the calculation of the atmospheric transmission coefficients, is strongly varying in space and time. Over terrestrial surfaces, it must therefore be measured or directly inferred from suitable scene data. In the operational approach presented here, τ_a is estimated by using clear lakes as reference target as proposed by Royer et al. (1988).

The accuracy of the method has been evaluated by comparisons between contemporary radiometric ground measurements and atmospherically corrected TM data from the years 1984, 1986 and 1988 (Hill & Sturm, 1989). A comparison of pseudo-invariant targets indicates also in this case, that the method provides an adequate solution for operational, i.e. scene-based, atmospheric corrections (figure 1-2).



5. Multi-temporal spectral features

The classification of multi-temporal data sets involves techniques for spectral feature extraction and data reduction. Greenness and brightness features (vegetation indices, Tasseled Cap features) have proven their suitability to emphasize important object characteristics for crop and vegetation monitoring, and they do provide a valuable reduction of data dimensionality.

A green leaf's chlorophyll pigment strongly absorbs sun radiation at wavelength between 0.5 and 0.70 μm (visible red) and the reflectance factor is normally below 0.1. In the near-infrared region (0.75 - 1.35 μm), multiple scattering occurs due to the leaf's internal mesophyll structure and the reflectance tends to be in the range of 0.4 to 0.6. This physiological relationship has been used to estimate the greenness of plant canopies through the use of various ratios (i.e. simple ratio NIR/R, normalized difference $(\text{NIR}-R)/(\text{NIR}+R)$). The normalized difference vegetation index

$$\text{NDVI} = (\rho_{tm4} - \rho_{tm3}) / (\rho_{tm4} + \rho_{tm3}) \quad (5.1)$$

is a bounded ratio between -1.0 and +1.0. Field observations indicate values between 0.1 and 0.8, which is confirmed by the analysis of atmospherically corrected TM data.

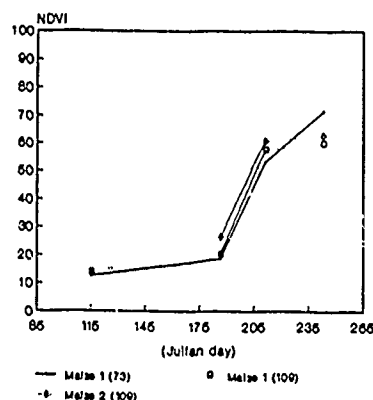
The Tasseled Cap transformation of Landsat TM data is a linear transformation of the form

$$\begin{bmatrix} \text{TC}_{11} & \text{TC}_{12} & \dots & \text{TC}_{16} \\ \text{TC}_{21} & \text{TC}_{22} & \dots & \text{TC}_{26} \\ \vdots & \vdots & \ddots & \vdots \\ \text{TC}_{61} & \text{TC}_{62} & \dots & \text{TC}_{66} \end{bmatrix} \times \begin{bmatrix} \text{DN}_1 \\ \text{DN}_2 \\ \vdots \\ \text{DN}_6 \end{bmatrix} = \begin{bmatrix} \text{F}_1 \\ \text{F}_2 \\ \vdots \\ \text{F}_6 \end{bmatrix} \quad (5.2)$$

with the coefficient matrix TC , the image signature vector DN and the vector F with the resulting features, the first three of which are known as brightness, greenness and wetness. Tasseled Cap transforms preserve euclidean relationships in the data space but capture typically 95 % or more of the total variability in its three features (Crist & Ciccone 1994). The transformation matrix is derived with respect to physical scene characteristics and is invariant to the spatial structure of individual scenes. It is therefore possible to compare quantitatively the derived features from different scenes, once the data have been normalized for external effects such as haze level, viewing and illumination geometry.

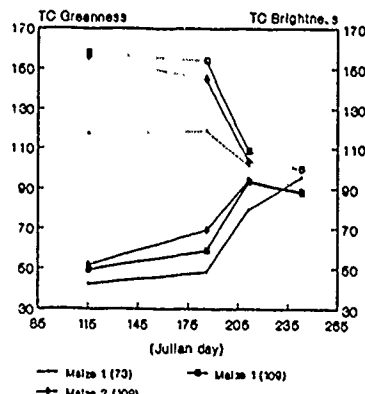
The conversion of the atmospherically corrected TM data into brightness, greenness and wetness was performed by using a coefficient

Figure 3:
Multi-temporal Vegetation Index (NDVI)
Maize, Segment #73/109 (Chomerac)



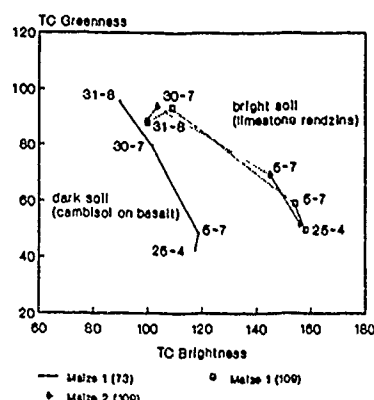
JRC Hls '89

Figure 4:
Multi-temporal TC Brightness & Greenness
Maize, Segment #73/109 (Chomerac)



JRC Hls '89

Figure 5:
Multi-temporal TC Brightness & Greenness
Maize, Segment #73/109 (Chomerac)



JRC Hls '89

matrix which has been developed for reflectance factor data (Crist, 1985).

In agricultural zones with a limited variety of crop types, good classification results have been achieved by using only the multi-temporal greenness index (Lo et al., 1986). There is, however, a distinctly different amount of original information retained in the NDVI and Tasseled Cap features. Greenness profiles as obtained from the NDVI, for example, indicate an almost identical development of three maize fields which belong to different ground observation segments (figure 3). The same time series of Tasseled Cap greenness reveals already a pronounced difference between the crops from segment 73 and 109, which apparently is related to the observed brightness level (figure 4). A check of the soil map showed that both ground observation segments are localized on different soil types. It is seen, that this brightness difference decreases with the progressing canopy closure of the maize crops (figure 5).

This gives an indication that the amount of information being preserved in the greenness feature alone, is not adequate for the multi-temporal analysis of more complex spatial entities with a large variety of natural vegetation types and agricultural crops. We use consequently all three Tasseled Cap features (brightness, greenness and wetness) for the multi-temporal classification, which still gives a data reduction of 50 %.

6 Multi-temporal classification

With respect to the relative small sampling fraction and some additional problems being related to the available ground observations (Hill & Megier, 1988), it was decided to pursue an unsupervised classification approach.

6.1 Image sampling and clustering

The unsupervised approach first involves the selection of candidate pixels from the 12-channel data set (brightness, greenness and wetness from 4 dates). This was initiated by using the multi-temporal NDVI for stratifying the test scene into permanently unvegetated (bare soil, water, urban etc.), permanently vegetated (forests, shrublands, permanent grassland) and areas of dynamically changing vegetation (agricultural

crops). Multi-temporal spectral signatures were extracted according to these stratification masks, where the sampling density could be adapted to the spatial extension and complexity of each stratum. Then euclidean distance-based disjoint cluster analysis of the spectral signature samples was employed to find local density maxima of the spectral groupings.

It is difficult to apply this technique in single-date image analysis. In the multi-temporal case, however, the thematic identification of clusters is greatly facilitated by the construction of temporal greenness profiles (figure 4). The examination of these profiles can be based on the knowledge about the local crop calendar and general characteristics of spectral responses (Lo et al. 1986). It is evident that reliable atmospheric corrections of the multi-temporal data have a great importance for this identification process.

Due to the complexity of the study region a high number of spectral classes is required to achieve optimal classifications. In the case study presented here, totally 54 sub-classes were mapped to the image domain, which were finally condensed to the 12 thematic classes of interest.

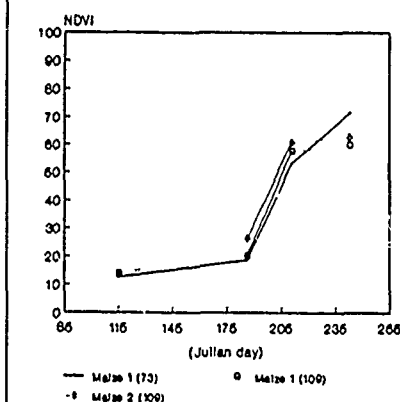
6.2 Spectral class mapping

It was on the basis of the minimum euclidean distance that all pixels of the multi-temporal data set were assigned to the land cover categories of importance. In this case, the distance d between a class center c_i and a pixel p is given by

$$d_{(c_i,p)} = \sqrt{(c_{i1}-p_1)^2 + (c_{i2}-p_2)^2 + \dots + (c_{in}-p_n)^2} \quad (6.1)$$

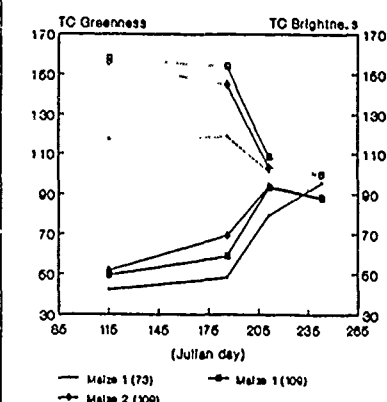
and p will be assigned to the class c_j if $d_j < d_i$ for all $i \neq j$, and $d_j < r$, otherwise p will be assigned to the rejection class c_r . The reject distance r is defined for each class according to its individual variance. This method was primarily chosen because of the high number of spectral classes. But it was assumed that, considering the increased discrimination potential of the multi-temporal data, such a simple distance measure would be adequate for successful class mapping.

Figure 3:
Multi-temporal Vegetation Index (NDVI)
Maize, Segment #73/109 (Chomerac)



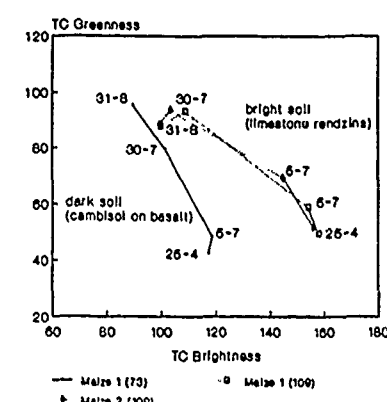
JRC HIN '89

Figure 4:
Multi-temporal TC Brightness & Greenness
Maize, Segment #73/109 (Chomerac)



JRC HIN '89

Figure 5:
Multi-temporal TC Brightness & Greenness
Maize, Segment #73/109 (Chomerac)



JRC HIN '89

matrix which has been developed for reflectance factor data (Crist, 1985).

In agricultural zones with a limited variety of crop types, good classification results have been achieved by using only the multi-temporal greenness index (Lo et al., 1986). There is, however, a distinctly different amount of original information retained in the NDVI and Tasseled Cap features. Greenness profiles as obtained from the NDVI, for example, indicate an almost identical development of three maize fields which belong to different ground observation segments (figure 3). The same time series of Tasseled Cap greenness reveals already a pronounced difference between the crops from segment 73 and 109, which apparently is related to the observed brightness level (figure 4). A check of the soil map showed that both ground observation segments are localized on different soil types. It is seen, that this brightness difference decreases with the progressing canopy closure of the maize crops (figure 5).

This gives an indication that the amount of information being preserved in the greenness feature alone, is not adequate for the multi-temporal analysis of more complex spatial entities with a large variety of natural vegetation types and agricultural crops. We use consequently all three Tasseled Cap features (brightness, greenness and wetness) for the multi-temporal classification, which still gives a data reduction of 50 %.

6 Multi-temporal classification

With respect to the relative small sampling fraction and some additional problems being related to the available ground observations (Hill & Megier, 1988), it was decided to pursue an unsupervised classification approach.

6.1 Image sampling and clustering

The unsupervised approach first involves the selection of candidate pixels from the 12-channel data set (brightness, greenness and wetness from 4 dates). This was initiated by using the multi-temporal NDVI for stratifying the test scene into permanently unvegetated (bare soil, water, urban etc.), permanently vegetated (forests, shrublands, permanent grassland) and areas of dynamically changing vegetation (agricultural

crops). Multi-temporal spectral signatures were extracted according to these stratification masks, where the sampling density could be adapted to the spatial extension and complexity of each stratum. Then euclidean distance-based disjoint cluster analysis of the spectral signature samples was employed to find local density maxima of the spectral groupings.

It is difficult to apply this technique in single-date image analysis. In the multi-temporal case, however, the thematic identification of clusters is greatly facilitated by the construction of temporal greenness profiles (figure 4). The examination of these profiles can be based on the knowledge about the local crop calendar and general characteristics of spectral responses (Lo et al. 1986). It is evident that reliable atmospheric corrections of the multi-temporal data have a great importance for this identification process.

Due to the complexity of the study region a high number of spectral classes is required to achieve optimal classifications. In the case study presented here, totally 54 sub-classes were mapped to the image domain, which were finally condensed to the 12 thematic classes of interest.

6.2 Spectral class mapping

It was on the basis of the minimum euclidean distance that all pixels of the multi-temporal data set were assigned to the land cover categories of importance. In this case, the distance d between a class center c_j and a pixel p is given by

$$d_{(c_j, p)} = \sqrt{(c_{j1} - p_1)^2 + (c_{j2} - p_2)^2 + \dots + (c_{jn} - p_n)^2} \quad (6.1)$$

and p will be assigned to the class c_j if $d_j < d_i$ for all $i \neq j$, and $d_j < r$, otherwise p will be assigned to the rejection class c_r . The reject distance r is defined for each class according to its individual variance. This method was primarily chosen because of the high number of spectral classes. But it was assumed that, considering the increased discrimination potential of the multi-temporal data, such a simple distance measure would be adequate for successful class mapping.

6.3 Classification results

The main scope of this case study is an assessment of the local mapping accuracy with multi-temporal data. Our preliminary evaluation, however, had to be restricted to an analysis of the 5 ground observation segments within our test scene which include mainly agricultural cover types (table 1 and 2). Given these restrictions, it can be stated that the multi-temporal approach gives highly satisfactory results. The overall accuracy is 79.1 %, which is the summation of the accuracy of each class weighted by the number of pixels occurring in each class (table 1). Classification errors are mostly related to the forest and shrubland classes, which is explained by the small extension of the respective areas within the agricultural land, and the inseparability of temporal signatures from very similar cover types.

Table 1. Error matrix comparing ground reference (seg 50, 51, 73, 109, 110) and unsupervised multi-temporal classification

| Class. | Ground Observations | | | | | | | |
|--------------|---------------------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 Unclassif. | 0 | 3 | 2 | — | 48 | 2 | — | 63 |
| 0 Bare Soil | — | 6 | 3 | — | — | — | 3 | 20 |
| 1 Dec Forest | 131 | — | — | — | 6 | 3 | 6 | 158 |
| 2 Grains | 4 | 207 | 5 | 1 | 17 | 1 | 0 | 240 |
| 3 Maize | 4 | — | 666 | 4 | — | — | 9 | 710 |
| 4 Sunflower | — | — | 6 | 72 | — | — | — | 79 |
| 5 Alfalfa | 2 | — | 5 | 3 | 311 | — | — | 322 |
| 6 Grasslands | 45 | 13 | 5 | 0 | 23 | 23 | 32 | 242 |
| 7 Shrublands | 20 | 5 | 5 | 1 | 17 | — | 95 | 155 |
| 8 Orchards | 25 | 3 | 26 | 11 | 3 | 2 | 3 | 344 |
| | 239 | 237 | 723 | 92 | 425 | 101 | 148 | 2332 |
| | 54.8 | 87.3 | 92.1 | 78.3 | 73.2 | 92.1 | 64.2 | 79.1 |

Table 2. Error matrix comparing ground reference (seg 50, 51, 73, 109, 110) and unsupervised single-date classification

| Class | Ground Observations | | | | | | | |
|--------------|---------------------|------|------|------|------|------|-----|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 Unclassif. | — | — | 22 | 1 | 11 | 2 | 1 | 37 |
| 0 Bare Soil | 1 | — | 135 | — | — | — | 3 | 152 |
| 1 Dec Forest | 132 | — | — | — | 16 | 3 | 3 | 177 |
| 2 Grains | 24 | 115 | 33 | 4 | 8 | 2 | 35 | 289 |
| 3 Maize | 4 | 21 | 409 | 28 | 10 | — | 3 | 566 |
| 4 Sunflower | 2 | — | 25 | 20 | 9 | — | 8 | 65 |
| 5 Alfalfa | 4 | — | 11 | 17 | 299 | 1 | 7 | 340 |
| 6 Grasslands | 50 | 25 | 27 | 16 | 67 | 60 | 23 | 348 |
| 7 Shrublands | 15 | 19 | 1 | 2 | 1 | 1 | 10 | 68 |
| 8 Orchards | 7 | 57 | 59 | 4 | 14 | 32 | 55 | 290 |
| | 239 | 237 | 723 | 92 | 425 | 101 | 148 | 2332 |
| | 55.2 | 48.5 | 55.6 | 21.7 | 70.4 | 59.4 | 6.8 | 47.5 |

Single-date classification was performed by using the same spectral sub-classes as in the multi-temporal case, except that only the Tasseled Cap features from the best date (5 July) were used. As it was expected, we found a reduced mapping accuracy and increased spectral confusion (table 2). But the difference between the two approaches is even better characterized by the fact that it was not possible to identify consistent spectral signatures for all agricultural cover types.

Multi-temporal thematic mapping of the natural vegetation types was examined in comparison with the existing aerial photographs and topographic maps. This evaluation indicates a good mapping accuracy for extended areas of deciduous forest, wood- and shrublands, and permanent grasslands. The acreage of coniferous forests is slightly

overestimated in areas of cast shadow, but the mapping seems to be more precise than in the case of single-date analysis. It is believed that, in mountainous terrain, distinct improvements will be only achieved by applying modelistic corrections of terrain illumination prior to the classification.

7. Conclusions

An operational approach to multi-temporal analysis is been presented which includes adequate methods for geometric and radiometric pre-processing of multi-temporal data sets.

Tasseled Cap features were found to provide a valuable reduction of the data volume while maintaining most of the relevant spectral information. The multi-temporal NDVI could be effectively used for a preparatory stratification into major cover types, but the lack of valuable intensity (general brightness) information limits its use for detailed classification in spatially complex and diverse landscapes.

It is demonstrated that multi-temporal TM data permit adequate classification and mapping of important land cover types even in areas of variable morphology and diverse agricultural and natural vegetation. Optimum multi-temporal classification must involve an increased number of spectral classes, but the computational requirements are partly compensated by the possibility of achieving precise results with less demanding algorithms.

8. References

- Crist, E.P. & R.C. Ciccone, "A physically-based transformation of Thematic Mapper data — the TM Tasseled Cap", *IEEE Trans. Geoscience & Remote Sensing*, vol. GE-22, no. 3, 256–263, 1984.
- Crist, E.P., "A TM Tasseled Cap equivalent transformation for reflectance factor data", *Remote Sensing of Environment*, 17, 301–306, 1985.
- Hill, J. & J. Mejean, "Regional land cover and agricultural area statistics and mapping of the Département Ardeche, France by use of Thematic Mapper data", *Int. J. Remote Sensing*, vol. 9, no. 9/10, 1573–1595, 1988.
- Hill, J. & B. Sturm, "Radiometric corrections of multi-temporal Thematic Mapper data for the use in agricultural land-cover mapping and vegetation monitoring", *Int. J. Remote Sensing*, acc. for publication, 1989.
- Kohl, H.G. & J. Hill, "Geometric registration of multi-temporal TM data over mountainous areas: by use of a low resolution digital elevation model", *Proc. 8th EARSeL Symp. on "Alpine and Mediterranean Areas: A Challenge for Remote Sensing"*, Capri (Naples), Italy, 17–20 May, 323–335, 1988.
- Lo, T.H.C., F.L. Scarpace & T.M. Lillesand, "Use of multi-temporal spectral profiles in agricultural land-cover classification", *Photogr. Engineering and Remote Sensing*, vol. 52, no. 4, 535–544.
- Royer, A., L. Charbonneau & P.M. Teillet, "Interannual Landsat-MSS reflectance variations in an urbanized temperate zone", *Remote Sensing of Environment*, 24, 423–446, 1988.
- Tanre, D., M. Herman, P.Y. Deschamps & A. de Lefre, "Atmospheric modeling of the background contribution upon space measurements of ground reflectance, including bi-directional properties", *Applied Optics*, 18, 3587–3594, 1979.
- Tanre, D., P.Y. Deschamps, P. Duhaet & M. Herman, "Adjacency effect produced by the atmospheric scattering in Thematic Mapper data", *Journal of Geophysical Research*, vol. 92, no. D10, 12000–12006.

LAND COVER ANALYSIS OF MULTITEMPORAL THEMATIC MAPPER DATA

S.Ueno

Information Science Laboratory, Kyoto School of Computer Science,
Sakyoku, Kyoto 606, Japan,

and

Y.Kawata and T.kusaka

Kanazawa Institute of Technology, Nonoichimachi, Ishikawa 921, Japan

ABSTRACT

In the present paper the time-sequential analysis of Landsat-5 Thematic Mapper (TM) data in Computer Compatible Tape (CCT) has been performed for diverse land cover types in Kanazawa area, Takayama district, in Japan. The subscene of TM data in the study (or test) site covering around 800X800 pixels of 25 meters square was extracted from the full scene data acquired on August 14, October 1, 1984, and August 1, 1985. After the preprocessing i.e., the radiative correction and geometric correction due to the reference points, on using the ground truth data covering a 20km square area with cartographic accuracy of 25m, we dealt with the supervised classification of both grey levels in TM and the principal components (PC) data acquired in 1984, and 1985. Furthermore, we performed the unsupervised classification of TM and PC data in 1984 and 1985. Finally, the statistical analysis of such quantities as the atmospheric reflectance, means of grey levels and others for both TM data in 1984 and 1985 has been done for the overall TM bands such that the multitemporal analysis of atmospheric reflectance permitted to make reasonable allowance for the comparative study of the above land cover analysis.

Keywords: Image Processing of Landsat Data, Radiometric and Geometric Corrections, Classification of Land-cover types.

1. INTRODUCTION

Since the launching of LANDSAT-4 and -5, the TM data for various kinds of land-cover types from the aspects of sensor performance, cartographic accuracy and image processing have been fully discussed by several authors (cf. Ref.(1)-(12)), allowing for the time-sequential and comparative study of TM and MSS data.

Our present paper aims to make the time-sequential analysis of an optimum information for the ecological study over Kanazawa area in Takayama district in Japan. The subscene of TM data in study (or test) site covering around 800X800 pixels was extracted from the above full scene data acquired on August 14 and October 1, 1984, and August 1, 1985. On using the ground truth data covering a 20km square area with cartographic accuracy of 25m span, we dealt with the supervised classification of both TM and PC data acquired in August and October. In addition to the above classification, we performed the unsupervised classification of TM and PC data in August and October. Finally, the statistical analysis of such quantities as the atmospheric reflectance, means of grey levels, and others for both TM

data in August and October has been done for the overall TM bands such that the multitemporal analysis of atmospheric reflectance permitted to make reasonable allowance for the seasonable change. As a whole it may be stated that the spectral classification of Landsat-5 TM data for various kinds of land-cover types in Kanazawa area has shown satisfactory results in accuracy.

2. PREPROCESSING

2.1 Preliminary data handling

2.1.1 Data acquisition: The Landsat-5 TM data in digital format handled by us have been acquired by the Earth Observation Center (EOC), National Space Development Agency of Japan (NASDA), and then have been preprocessed in terms of the radiometric and geometric calibration in systematic sense by EOC.

According to the annotation records, the handling data are listed in Table 1. The TM data scene under consideration produced as a P-tape (CCT-PT) by EOC covered the land-use map (1:25,000) of such areas as "Kanaiwa", "Awagasaki", "Matsuto", and "Kanazawa", in Ishikawa prefecture, Japan.

Table 1. Constants of Acquired TM Data

| | Acquired Data in 1984 | | Acquired Data in 1985 |
|-------------------|-----------------------|-------------------|-----------------------|
| | August 14 | October 1 | August 1 |
| Path-row | 109-35 | 109-35 | 109-35 |
| Central Latitude | 136°54' | 136°57' | 137°1' |
| Central Longitude | 36°04' | 36°03' | 36°03' |
| Solar Altitude | 55° | 42° | 57° |
| Solar Azimuth | 121° | 143° | 116° |
| ID Number | E-50166-00575-0 | E-50214-00581 | E-50181-00575-0 |
| Resampling Scheme | Cubic Convolution | Cubic Convolution | Cubic Convolution |

2.1.2 Data processing: Ground cover conditions in the above areas were identified by a joint work of Geographical Survey Institute (GSI) and Ishikawa Research Laboratory for Public Health and Environment (IRLPHE), by making use of the land use maps, aerial photographs, and meteorological data. The total number of ground truth data in pixels is a few thousands in our study site. The cartographic accuracy was shown to be within twenty five meters in North-South and East-

[illegible]

Table 5. Classification accuracy (%) of TM data on October 1, 1984

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|-----|------|
| 1 | 41.7 | 8.3 | 0.0 | 33.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 16.7 | 0.0 | 0.0 |
| 2 | 19.0 | 47.6 | 0.0 | 0.0 | 23.8 | 0.0 | 0.0 | 9.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 10.0 | 0.0 | 50.0 | 20.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 20.0 | 0.0 | 0.0 |
| 4 | 8.2 | 0.0 | 0.0 | 93.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 25.0 | 0.0 | 0.0 | 3.8 | 71.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 6 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 | 75.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 7 | 0.0 | 5.1 | 0.0 | 0.0 | 33.3 | 0.0 | 82.1 | 12.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 11.3 | 80.8 | 8.1 | 0.0 | 0.0 | 0.0 | 0.0 | 12.5 |
| 9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 87.2 | 0.0 | 0.0 | 0.0 | 0.0 | 2.8 |
| 10 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 87.5 | 12.5 | 0.0 | 0.0 | 0.0 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 77.8 | 0.0 | 0.0 | 22.2 |
| 12 | 12.5 | 12.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 75.0 | 0.0 | 0.0 |
| 13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100 | 0.0 |
| 14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 93.8 |

Table 6. Classification accuracy (%) of TM data on August 1, 1984

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----|------|------|------|------|------|------|------|------|------|------|------|-----|
| 1 | 58.2 | 32.2 | 0.0 | 0.0 | 0.0 | 4.4 | 3.7 | 0.0 | 5.2 | 0.0 | 0.0 | 0.0 |
| 2 | 19.0 | 81.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 0.0 | 0.0 | 94.7 | 0.0 | 5.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 0.0 | 0.0 | 0.0 | 82.5 | 0.0 | 7.2 | 10.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 0.0 | 13.1 | 0.0 | 75.3 | 0.0 | 11.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 80.3 | 19.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 7 | 0.0 | 0.0 | 0.0 | 4.0 | 0.0 | 19.0 | 77.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 74.1 | 25.9 | 0.0 | 0.0 | 0.0 |
| 9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.8 | 0.0 | 7.3 | 89.1 | 0.0 | 0.0 | 0.0 |
| 10 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 90.2 | 9.8 | 0.0 |
| 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.0 | 82.0 | 0.0 |
| 12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100 |

3.2 Unsupervised classification

In a manner similar in form to the clustering analysis (cf. Ref.(10)), the unsupervised classification was performed by the grey levels in CCT and PC data in August and October, 1984 and 1985, such that, on making use of the transformed divergence of the clustering separability, the perfield classification was implemented. Compared with the landuse map printed in colours, the image consisting of fifteen classes on August 1 1985, obtained by the clustering analysis for PC data showed such that the coastal line in Japan sea appeared clearly without noise and furthermore the particular features of the ridge between the rice fields were noticeable. The comparison of the supervised classification with the unsupervised ones showed that, from the aspect of the efficiency and performance the latter method has seemed to be a little superior to the former ones, particularly in the sense that the peculiar patterns in the rice field obtained by the clustering method were noticeable compared with that classified by the GML method.

4. SPECTRAL STATISTICS OF TM DATA

The statistical analysis of spectral radiance in digital counts was performed, in virtue of which the spectral characteristics of TM data for several cover types in study (and test) site become clearer. For this purpose the various cover types listed in Table 3 were further condensed in four categories (Ref.(11)). The radiance counts extracted from these sample fields were used to get the spectral radiance values L_b by Eq.(1), means, variance, coefficients of variation (variance/means) $\times 100\%$, minimum, and maximum values, correlation matrix, covariance matrix for the given seven TM bands in main cover types. These statistical summaries for the aqueous, structural, ecological and other study sites were listed in tables, respectively, except for the spectral radiance and covariance matrix (cf. Ref.(11)).

It has been discussed in our preceding paper

(Ref.(11)) that bands 5 and 7 in the aqueous study site showed a remarkable coefficient of variance resulting from highly variable reflectance phenomena, whereas band 6 was the least variable band due to the low magnitude and narrow range of radiation. Furthermore, the atmospheric (or relative) reflectance was evaluated as below,

$$I = L_b / (uF) \quad (2)$$

where L_b is the spectral radiance, $\cos^{-1}u$ is the polar angle of the sun, and F is the solar spectral irradiance (or flux) above the atmosphere on a unit area normal to the direction of propagation. The solar flux F in each spectral band is listed in Table 7. Atmospheric reflectance for representative land cover-types versus each TM bands is plotted in Figure 1, on August 14, 1984, in Figure 2, on October 1, 1984, in Figure 3 on August 1, 1985, respectively. It was shown that most remarkable change of the atmospheric reflectance between August and October was recognized in the ecological cover-types in study area. It is understood that the change of the atmospheric reflectance between summer and fall seasons became remarkable in ecological area, such as in the rice field, field, meadow, and forest. On the other hand, we did not recognize any significant change in atmospheric reflectance in the aqueous, structural, and other areas in August and October.

Table 7. Solar Flux F ($\text{mW}\cdot\text{cm}^{-2}$) in Spectral Band of TM Data

| Band | Wave-length | F | Band | Wave-length | F |
|------|-------------|-------|------|-------------|-------|
| 1 | 0.45-0.52 | 13.65 | 4 | 0.76-0.90 | 15.06 |
| 2 | 0.52-0.60 | 14.65 | 5 | 1.55-1.75 | 4.64 |
| 3 | 0.63-0.69 | 9.24 | 7 | 2.08-2.35 | 2.16 |

5. DISCUSSIONS

In the present paper, making use of Landsat-5 TM CCT data acquired on August 14 and October 1, 1984, and August 1, 1985, we performed an image processing of the subscene consisting of the 800X800 pixels of 25 meters square size in Kanazawa district, Japan. After the preprocessing consisting of the radiometric and geometric (GCP) corrections, our TM CCT data in August and October have been classified by means of the GML classifier. Whereas the ground truth data in magnetic tape given by GIS and IPLPHE have been classified in forty nine categories, however, we summarized it in fourteen (or twelve) categories because of the somewhat redundant classification scheme of the former based on the administrative aspects. The classification accuracy in fourteen (or twelve) land cover types is listed in percentage in Tables 4, 5 and 6. The integer numbers in first row and first column represent the legend number in Table 3.

Furthermore, with the aid of the principal component (PC) analysis, we obtained the PC first, second, and third channels for TM CCT data on August 14, 1984, and August 1, 1985, and classified them with the aid of the GML classifier and the clustering analysis. The comparative study of the classification results for TM CCT data and PC data showed us such that there was no distinguished difference between them, whereas the machine time of the classification procedure for PC data was much less than that for TM CCT data.

Because of somewhat short term of the TM CCT data acquired in August and October, 1984, and in August, 1985, the remarkable changes in classification results

have not been found except for the ecological land-cover type. The classification accuracy of TM CCT data was proved to be satisfactory. Furthermore, the statistical analysis of the grey levels of the emergent radiance, i.e., means, variance, and others has shown an interaction relationship of these TM data in aqueous, structural, ecological, and other areas in each spectral band.

Finally, the atmospheric reflectance in terms of the emergent radiance, solar polar angle, and solar flux has been plotted against for each spectral band in four main categories of the land cover types. It was found that the conspicuous difference of the zigzag line slope has been recognized only in ecological category of land-cover type in August and October.

ACKNOWLEDGEMENTS

The authors' sincere appreciation is extended to Y.Hoshino and H.Masaharu in GSI and also to T.Tajima and Y.Takeno in IPLPHE who offered us the magnetic tape of ground truth data in Kanazawa area. We also are grateful for support under contract No.60129032 of the research grant of the Japanese Ministry of Education.

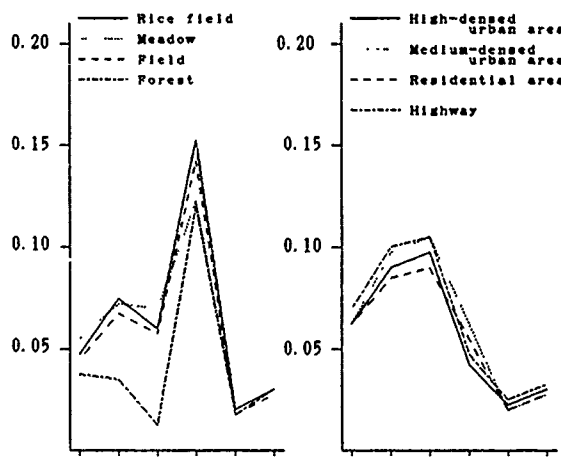


Fig. 1 Atmospheric reflectance versus each of TM bands for the ecological (left-hand side) and structural (right-hand side) areas on August 14, 1984.

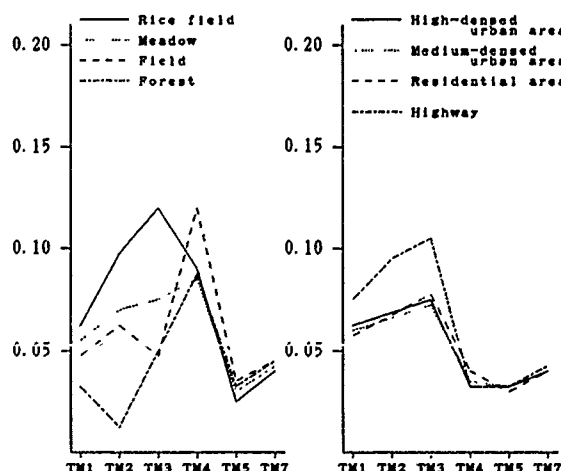


Fig. 2 Atmospheric reflectance versus each of TM bands for the ecological (left-hand side) and structural (right-hand side) areas on October 1, 1984.

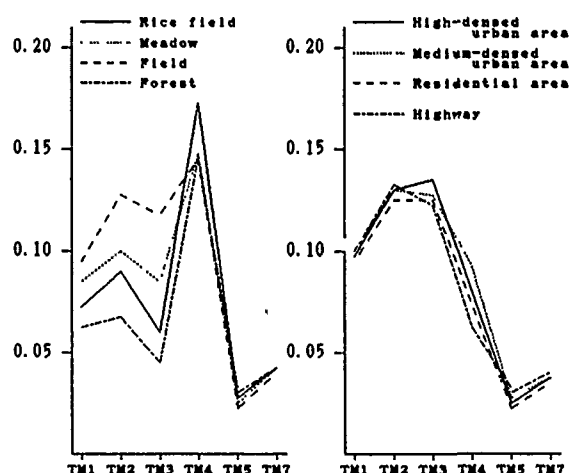


Fig. 3 Atmospheric reflectance versus each of TM bands for the ecological (left-hand side) and structural (right-hand side) areas on August 1, 1985.

REFERENCES

- Bernstein, R., Intsplech, J.B., Meyers, H.J., Kolsky, H.G., and Lees, R.D., 1984. "Analysis and processing of LANDSAT-4 sensor data using advanced image processing techniques and technologies", IEEE Transaction of Geoscience and Remote Sensing, Vol. GE-22, pp.192-221.
- Duggin, M.J., Rowntree, R., Emmons, M., Hubbard, N., Odell, A.W., Sakhavat, H., and Lindsay, J., 1986. "The use of multitemporal multichannel radiance data in urban feature analysis", Remote Sensing of Environment, Vol.20, pp.95-105.
- Hill, J., and Sturm, B., 1988. "Image-based atmospheric correction of multitemporal Thematic Mapper data for agricultural land cover classification", in Proc. of IGARSS'88 Symp. Edinburgh, Scotland, 13-16 Sept. 1988, pp.895-899.
- Jiaju, Lu., 1988. "Development of principal component analysis applied to multitemporal Landsat TM data", Int. J. Remote Sensing, Vol.9, pp.1895-1908.
- Kalayeh, H.M., and Landsgrebe, D.A., 1986. "Utilizing multitemporal data by a stochastic model", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-24, pp.792-795.
- Kavata, Y., Kusaka, T., and Ueno, S., 1983. "A fundamental approach to temporal data analysis", in Proc. of the 16th Int. Symposium on Remote Sensing of Environment, Buenos Aires, Argentina, June, 1983, pp.533-544.
- Kusaka, T., Kavata, Y., and Ueno, S., 1984. "Multitemporal data analysis by extended radiometric correction", in Proc. of the 18th International Symposium on Remote sensing of Environment, 1-5 October, 1984, Paris, France, pp.1775-1783.
- Lynn, D.W., 1986. "Monotemporal, multitemporal, and multirate thermal infrared data acquisition from satellites for soil and surface-material survey", Int. J. Remote Sensing, Vol.7, pp.213-232.
- NASA Technical Memorandum 86149, 1984. "A prospectus for thematic mapper research in the earth sciences".
- Reeves, R.G., (ed.) 1975. "Manual of Remote Sensing", Vol.I,II, American Society of Photogrammetry, Falls Church, VA, USA.
- Ueno, S., Kavata, Y., and Kusaka, T., 1985. "Optimum classification of LANDSAT Thematic Mapper data for ecological study", in Proc. of the 19th International Symposium on Remote Sensing of Environment, October 21-25, 1985, Ann Arbor, Michigan, pp.533-544.
- Ueno, S., Kusaka, T., and Kavata, Y., 1986. "Bitemporal analysis of Thematic Mapper data for land cover classification", in Proceedings of IGARSS'86 Symposium, Zurich, 8-11 Sept, 1986, Ref. ESA SP-254, pp.523-528.

MICROWAVE REMOTE SENSING OF LAND SURFACES WITH AIRBORNE

ACTIVE AND PASSIVE SENSORS

P. Coppo *, P. Ferrazzoli **, G. Luzi *, P. Pampaloni ***, G. Schiavon **, D. Solimini **

* Centre for Microwave Remote Sensing, Viale Galileo 32- 50125 Florence, Italy

** Dip. Ingegneria Elettronica "Università Tor Vergata", Via O. Raimondo- 00173 Rome, Italy

TX 611462 Fax 39-6-2490519

*** IROE - CNR, Via Panciatichi 64 - 50127 Florence, Italy

TX 570231 Fax 39-55-410893

Abstract

Observations of microwave emission and backscattering from soil and vegetation were carried out by means of airborne sensors during Agrisar 86 and Agriscatt 87 campaigns. The experiments aimed at clarifying some relations between sensor responses and ground parameters and at identifying the dominant features which affect emission and scattering. Some examples of these relations are shown in this paper. Besides, a comparison between emissivity and backscattering is carried out with the purpose of checking the consistency of measurements.

Keywords: Microwave, emission, scattering, vegetation.

Introduction

The dependence of microwave emission and scattering from terrain on ground parameters has been investigated in several experimental and theoretical works (Ulaby et al. 1981, 1982 and 1986). However, especially at frequencies higher than 5 GHz, the knowledge of vegetation microwave features is far from being exhaustive. The purpose of this paper is to show and discuss some experimental relations between remote sensing measurements and ground parameters by using data collected with active and passive microwave sensors.

Besides being synergistic in several practical applications, and although emission and scattering are biunivocally related only in particular cases, simultaneous observations of emissivity and backscattering can be very fruitful in the understanding of the global electromagnetic properties of natural media and for checking experimental data (Ferrazzoli et al. 1988). The analysis of the sensitivity of backscatter and emissivity to geophysical parameters is here carried out separately for the two quantities. However in some cases the two parameters have been directly compared with the main purpose of checking the consistency of measurements on the basis of a theoretical model derived from the energy conservation law.

Experiments

In 1986, the Joint Research Center of EEC organized an airborne backscatter measurement campaign, named AGRISAR which took place over several European countries. The Italian test site Oltrepò Pavese, located in Northern Italy, was surveyed four times in June, July and August by a B17 aircraft carrying the French Varan-S SAR, operating at 9.37 GHz (VV, HH polarizations) with nominal ground and radiometric

resolution of 3 m and 1 dB respectively (Vaillant 1985). A few hours after the second and third SAR flights (i.e. June 26th, July 16th) the site was surveyed by a small aircraft on which were mounted three radiometers (operating at X, Ka and infrared bands) backward looking at soil at $\theta \approx 40^\circ$ incidence angle (Bonsignori et al., 1987). The radiometer flights were organized by IROE-CNR and the Centre for Microwave Remote Sensing in Florence. During the last three SAR flights the most relevant ground parameters were measured.

The same site was surveyed 4 times in June and July 1987 by the Dutsat scatterometer, operating at six frequencies (L, S, C, X, Ku1 and Ku2 bands), four incidence angles (10, 25, 40, 55 degrees) and two polarizations (VV and HH). Ground resolution at X and Ku bands was less than 10 m (Snoeijs et al., 1987). The 1987 campaign was jointly organized by the European Space Agency and JRC. As in 1986 almost contemporary airborne radiometric measurements were carried out and ground truth data were collected; corn, sugarbeet, alfalfa, wheat and bare soil fields were routinely observed and some measurements over barley and stubbles were also performed. Ground truth measurements concerned the most important parameters of soil and vegetation, like soil moisture, plant height, density and water content, Leaf Area Index, etc.

The period covered by the flights (June-August) allowed the observation of a limited part of the crop growth cycles, particularly in the case of wheat and sugarbeet.

Results

The basic sensed parameters are the normalized radar cross section σ^0 (m^2/m^2) and the emissivity ϵ ; the latter, which can be rigorously defined only for homogenous isothermal bodies, is approximated here by a normalized temperature T_N , obtained by the ratio between microwave and infrared brightness temperatures.

Backscattering data at 9.4 GHz were obtained in 1986 from the Varan-S SAR data, by averaging the power levels of all the pixels ($\approx 3\text{m} \times 3\text{m}$) included in the fields which were observed at $\theta \approx 45^\circ$ incidence angle.

The 1987 scatterometric data at X band were found to be affected by instabilities and until now it has been practically impossible to work out correlations with land features. Ku band data are more reliable although the absolute values appear to differ (6-10 dB higher) from data available in literature. Since we expect that the behaviour of vegetation at 9.5 and 13.7 GHz is fairly similar, in this paper we examine

data at 13 GHz; σ^0 represents again the average value of all the measurements available in the same field. A relative scale is used because of the mentioned difficulties in calibration.

As far as the passive measurements are concerned the microwave normalized temperature at X band of each field has been obtained by considering all the samples collected in the central part of the field.

As expected the results of the measurements show that in general dry bare (or stubbles covered) soil has higher emissivity and lower backscatter than vegetated fields. However one has to remember that emission and scattering from bare soil strongly depend on surface moisture and roughness.

Among crops, wide leaf plants such as corn and sugarbeet exhibit higher σ^0 and lower brightness than small grains (wheat and barley) and alfalfa. However, it has already been shown (Ulabi et al., 1986; Pampaloni and Paloscia, 1985) that the remote sensing parameters depend not only on crop type but on vegetation conditions as well. The latter can be identified by means of appropriate agronomical parameters such as leaf area index -LAI- (in m^2/m^2), plant water content -PWC- (in kg/m^2 or in kg/m^3), plant moisture -M- (in % of H_2O on total weight) which are also indicators of the mass of water overhanging the soil and then of the dielectric characteristics of the air vegetation mixture.

The dependence of microwave data on the various ground parameters has been examined and some examples are reported as follows.

Fig. 1 shows the backscattering coefficient σ^0 at 9.4 GHz, measured on different crops, as a function of PWC in kg/m^2 ; we see that, although data are fairly highly spread, in general σ^0 increases as PWC increases, however ripe wheat (W) exhibits an opposite trend.

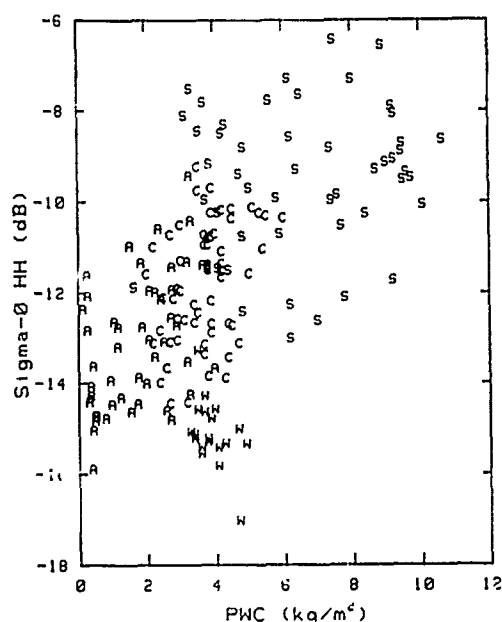


Fig.1: Backscattering coefficient obtained in 1986 (26th June, 16th July and 12th August) from VARAN-S data at HH pol., $\theta=45^\circ$, as a function of Plant Water Content (PWC) for different crops: A=alfalfa, B=barley, C=corn, S=sugarbeet, W=wheat

This behaviour is confirmed by the scatterometric data which also confirm the similar microwave behaviour of wheat and barley (B) (fig.2).

In turn radiometric measurements (fig.3) show a behaviour which is consistent with that of radar data. Here the general trend is a decrease of the normalized temperature as PWC increases; this global behaviour results from the combination of data for the various crops, whereas for each crop the sensitivity to PWC is quite low.

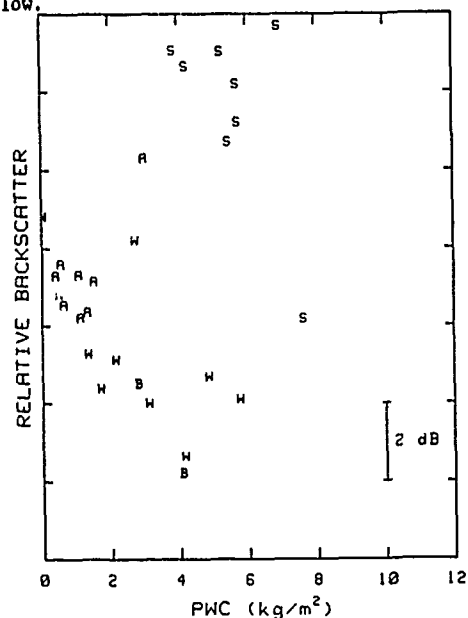


Fig.2: Relative backscatter, measured in 1987 (2nd and 17th June and 7th July) by scatterometer at $f=13.7$ GHz, HH pol., $\theta=40^\circ$ as a function of PWC for different crops; crop code is the same as fig. 1

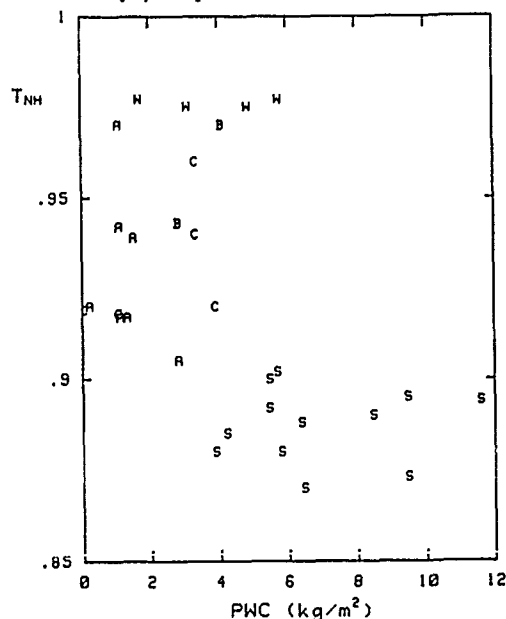


Fig. 3: Normalized brightness temperature, (X band, H pol., $\theta=40^\circ$) measured in 1986 (16th July) and 1987 (3rd and 17th June) on various crops, as a function of PWC; crop code is the same as fig. 1

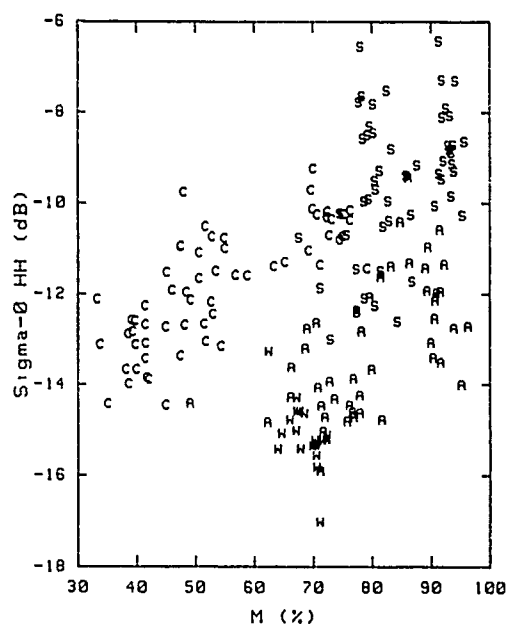


Fig. 4: Backscattering coefficient (X band, HH pol., $\theta=45^\circ$) as a function of % (by weight) of water in the plants (M). Observations refer to the senescence cycle of vegetation in 1986 (26th June, 16th July and 12th August); crop code is the same as fig. 1

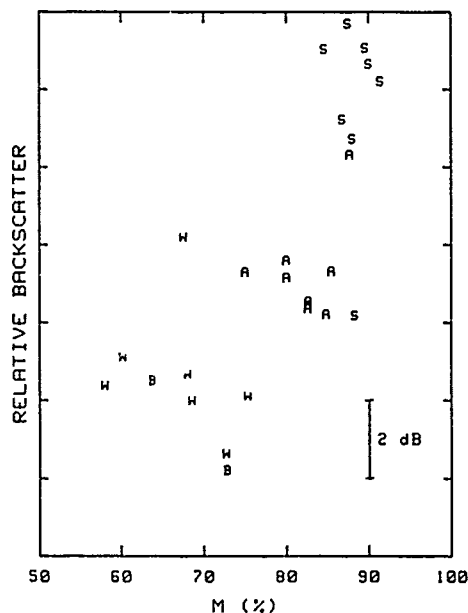


Fig. 5: Relative backscatter at 13.7 GHz, HH pol., $\theta=40^\circ$, as a function of M. Observations with Dutsat in June-July 1987; crop code is the same as fig. 1

The most evident feature exhibited by the brightness data is a stable difference between normalized temperatures of wheat and sugarbeet, which fall at opposite extremes of the range, whilst corn and alfalfa have intermediate values depending on the growth stage.

Once the crop growth is completed the dry biomass increases at a low rate and vegetation begins to lose water; so it appears significant to correlate the microwave parameters to the percentage of water in the plants, M. Diagrams of figures 4, 5 and 6 show that σ_{HH} increases and TN decreases as a function of M. We remind that for these measurements the variation of M is due to the vegetation senescence. In these diagrams for a given M we can recognize two ranges of σ° and e: the one corresponding to wide leaf crops (corn and sugarbeet) and that of alfalfa, barley and wheat. For the latter crops the sensitivity of microwave parameters to M seems to be slightly higher than for wide leaves.

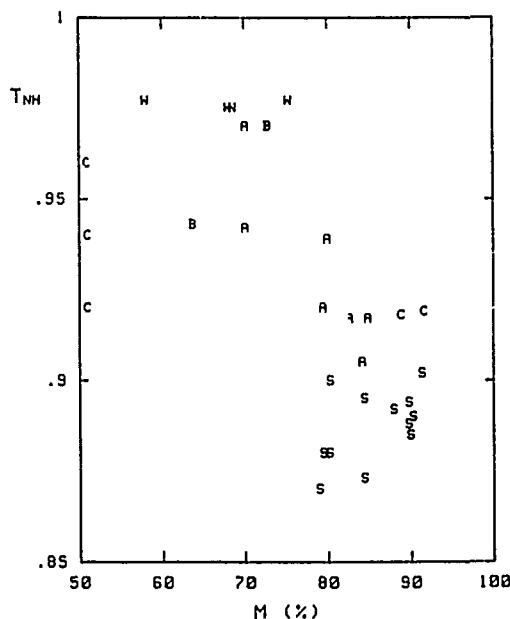


Fig. 6: Normalized brightness temperature (X band, H pol., $\theta=40^\circ$) as a function of M. Observations in 1986 (16th July) and 1987 (3rd and 17th June); crop code is the same as fig. 1

The previous diagrams show that in general, the backscattering coefficient increases as the normalized temperature decreases although general a direct relation between these two parameters does not exist.

However the problem of correlating emission and backscattering has been considered by Tsang et al. (1982) in the particular case of a uniform continuous medium (soil) covered by a layer of isotropic scatterers (leaves). Although this model may be far from most real conditions it can be used as a basis for discussing our results.

The equation found by Tsang et al. (1982) is the following:

$$\sigma_{\beta,\beta}(\theta) = [1 - e_{\beta}(\theta) - r_{\beta}^s(\theta) \exp(-2\tau \sec \theta)] \frac{\cos^2 \theta}{F_2(\tau, -\cos \theta)} * (1) \\ * [1 - \exp(-2\tau \sec \theta)] + \sigma_{\beta,\beta}^s(\theta) \exp(-2\tau \sec \theta)$$

where $\sigma_{\beta,\beta}(\theta)$ is the backscattering coefficient, $r_{\beta}^s(\theta)$ is the ground reflectivity, $e_{\beta}(\theta)$ is the emissivity, β is an index of the polarization, θ is the incidence angle, $\sigma_{\beta,\beta}^s(\theta)$ is the ground backscattering coefficient, τ is the optical depth, and $F_2(\tau, -\cos \theta)$ is a function described by Chandrasekhar (1960).

The model can be parameterized to find the conditions which best approximate experimental data. When vegetation is well developed and for observations far from nadir ($\geq 40^\circ$) the soil backscattering σ_{soil} may be neglected and the number of parameters is reduced. Diagram of fig. 7 represents experimental points, collected on alfalfa, corn and sugarbeet with different LAI values, compared with the model for a moistened soil having a reflectivity equal to 0.25 and optical depth τ ranging from 0.6 to 1.8. We see that the model reproduces fairly well the growth of alfalfa in that points corresponding to LAI = 1.4 (A1) and 4.3 (A4) lie on curves with consistent values of τ (Paloscia and Pampaloni, 1988). However the behaviour appears less consistent for corn, where only the point with LAI = 3.8 (C4) seems to be justifiable, and above all for sugarbeet, whose representative points would correspond to too low values of optical depth.

This result appears quite reasonable because at 3 cm wavelength alfalfa crop approximates fairly well a medium with isotropic scatterers (Ferrazzoli et al., 1988) whereas we may expect that crops with wide leaves such as corn and sugarbeet are far from that approximation. This is particularly true for dry soil covered with wheat stubbles or cut alfalfa stems; points representing measurements on these surfaces, where reflectivity is usually much lower than 0.20, lie in the region of relatively low τ . However a comparison of these data with the model is of little significance because eq. 1 is not able to give a valid relation between e and σ° for very low values of τ .

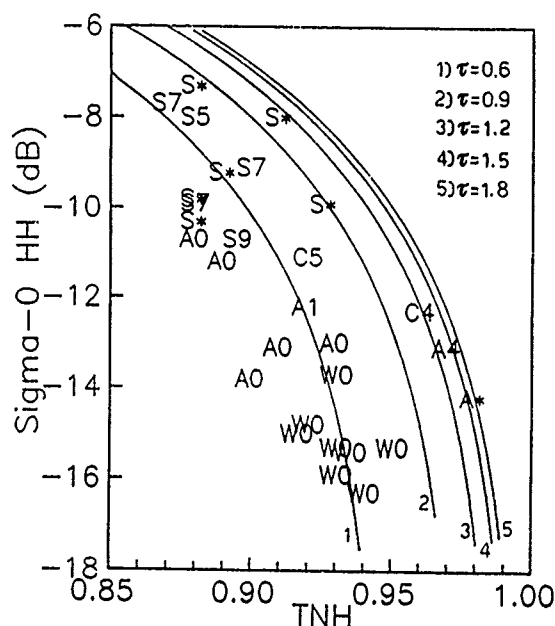


Fig. 7: Backscattering coefficient at X band, HH pol., $\theta=45^\circ$ as a function of the emissivity. Lines refer to the model of eq. 1 for a soil reflectivity equal to 0.25 and τ ranging from 0.6 to 1.8. The points, denoted by letters and numbers, represent measurements of backscattering and normalized temperature. A=alfalfa, C=corn, S=sugarbeet, W=wheat. Numbers represent rounded values of leaf area index, * denotes well developed crop with unknown leaf area index.

Conclusions

The analyzed experimental results show that microwave emission and backscattering at X band, H polarization, are sensitive to crop type and plant water conditions. Despite the spread of data a recurrent different behaviour between wide leaf crops and small grains can be identified. The latter crops are fairly well represented by a layer of isotropic scatterers whereas corn and sugarbeet plants must probably be modelled with a different approach. The dependence of emissivity and backscattering coefficient from plant water content is well established for the ensemble of crops during their life cycle, however the sensitivity of both parameters appears quite low for detecting plant water conditions of a single crop in a small time interval. A direct comparison of active and passive measurements, with the help of a theoretical model, makes possible a check of data and may allow a better crop separation than the one achievable by using only one sensor.

Acknowledgements

The work was partially supported by Italian Space Agency, formerly CNR-PSN and by M.P.I. The Center for Microwave Remote Sensing wish to thank Officine Galileo S.p.a. and SMA S.p.a. for their support.

REFERENCES

1. Bonsignori R., L. Chiarantini, S. Paloscia, P. Pampaloni, P. Ferrazzoli, M. Mongiardo, D. Solimini, "Aircraft microwave radiometry of land", Proc. of IGARSS'87, vol. II, THA-7, pp. 1429, 1987.
2. Chandrasekhar S., "Radiative transfer", New York 1960, Dover Publications, Inc., 1960.
3. Ferrazzoli P., G. Schiavon, D. Solimini, G. Luzi, P. Pampaloni and S. Paloscia, "Comparison between microwave emissivity and backscattering coefficient of agricultural fields", Proc. of IGARSS'88, Symp. Edinburgh, Scotland, ESA SP-284, pp. 667, 1988.
4. Paloscia S. and P. Pampaloni, "Microwave polarization index for monitoring vegetation growth", IEEE Trans. on Geoscience and Remote Sensing, vol. GE-26, n. 5, pp. 617-621, 1988.
5. Pampaloni P. and S. Paloscia, "Experimental relationships between microwave emission and vegetation features", Int. J. Remote Sensing, vol. 6, n. 2, pp. 315-323, 1985.
6. Snoeij P., P.J.F. Swart, The DUT airborne scatterometer, Int. J. Remote Sensing, vol. 8, n.11, pp. 1709-1716, (1987)
7. Tsang L., A. Blanchard, R.W. Newton and J.A. Kong, "A simple relation between active and passive microwave remote sensing measurements of earth terrain", IEEE Trans. on Geosci. and Remote Sensing, vol. GE-20 n. 4, pp. 482-485, 1982.
8. Ulaby F.T., R.K. Moore and A.K. Fung, "Microwave remote sensing", Vol. I,II,III, Artech House Inc., Dedham, Massachusetts, (1981, 1982 and 1986)
9. Ulaby F.T., M. Ratzani and M.C. Dobson, "Effects of vegetation cover on the microwave radiometer sensitivity to soil moisture", IEEE Trans. Geosci. and remote sensing, GE-21, n. 1, pp. 51-61, 1983.
10. Vaillant D., "VARAN-S: An airborne synthetic aperture radar for research in microwave remote sensing", Proc. EARSeL Symposium, Strasbourg, 1985.

IDENTIFICATION AND ASSESSMENT OF ANOMALIES ON RADAR IMAGERY OF CROPS

Glenn H. Holder¹ * Philip J. Howarth¹ Richard Mussakowski²

¹ Earth-Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (* Current address: TYDAC Technologies Inc., 1600 Carling Avenue, Suite 310, Ottawa, Ontario K1Z 8R7)

² Ontario Centre for Remote Sensing, Ontario Ministry of Natural Resources, 90 Sheppard Avenue East, 4th Floor, North York, Ontario M2N 3A1, Canada

Abstract

In remote sensing studies of agricultural crop characteristics, it is important to obtain ground information close to the date of acquisition of the imagery. This is particularly relevant when radar imagery is being analysed as it is often difficult to determine the ground conditions that influence the tonal appearance of the image. An Anomaly Assessment System has been developed for combining quicklook imagery, a digital field boundary map and ground data to rapidly identify those fields which have either lighter or darker tones than the average. Use of the system will permit rapid identification and field checking of anomalous fields.

Keywords: Anomaly assessment, Crop inspection, Crop inventory, Radar imagery

Introduction

In the majority of studies involving interpretation of remote sensing data, it is important to have ground information to aid in the analysis of the imagery. How detailed the ground information should be and how close it should be obtained to the date and the time of acquisition of the imagery vary depending upon the nature of the study being undertaken and the region of the electromagnetic spectrum in which the data are obtained. In this paper, we consider the acquisition requirements for crop identification and assessment, and suggest ways in which timely information can be obtained to aid in the analysis of radar imagery of crops.

Crop characteristics change during the growing season. Sometimes, these changes can take place over a period of just a few days (e.g., ripening of grain crops). Thus, it is important to make field observations fairly close to the date of image acquisition if one is going to accurately interpret the crop type and its stage of development. Working in the visible part of the spectrum, the eye observes the same wavelengths as the sensor. As a result, comparison of field conditions with the imagery is relatively easy. Moving into the near-infrared regions, however, some extrapolation between the field appearance of the crops and how they are recorded on the imagery has to be made. Within

the radar wavelengths, two sets of factors will influence the appearance of a specific crop type on the imagery. First are the characteristics of the radar system itself and its geometric position relative to the area of interest. Second, ground conditions will affect the radar backscatter and ultimately the tone or grey-level that is observed on the imagery. Ground conditions in agricultural areas display not only the influence of soil characteristics, such as the surface roughness and moisture content, but also the effects of crop conditions such as the stage of growth, the geometric arrangement and the moisture content of the crop itself (Ulaby and Bush, 1976; Ulaby and Bativala, 1976).

Although similar crops are planted in different fields throughout a region, ground conditions will vary depending upon factors such as soil type, organic content of the soil, slope and moisture content. Even with similar ground conditions, crop appearance will vary depending upon the time of planting, crop variety and the availability of fertilizer in the soil. Given the sensitivity of radar imagery to a wide range of environmental factors, it is often difficult to explain the differing appearances of the same crop type on an image (Hirose *et al.*, 1983; Cihlar, 1986). One way to attempt to do this is through the process of anomaly assessment. Often it is the anomalous appearance of just a few fields that can give us an insight into some of the factors that have a major influence on the appearance of the image. Thus, methods are required to rapidly identify anomalous fields for more detailed ground study as soon as possible after acquisition of the imagery.

Methodology

The sequence of analysis procedures for anomaly assessment is shown in Figure 1. The first stage is acquisition of radar imagery. During a SAR-580 overflight by the Canada Centre for Remote Sensing (CCRS), the radar data are recorded as quicklook imagery and as digital data. As processing of the digital data normally takes several months, it is the quicklook imagery which must be used for anomaly assessment. This imagery can be sent to the user in approximately 24 hours. On the quicklook imagery, the analyst is faced with a wide range of different

image tones displayed in a large number of fields. The problem is how to rapidly identify those fields that for one reason or another have image tones which are either distinctly lighter or darker than the average tone for the set of fields containing the same crop type. Clearly, it has to be done in an automated fashion.

To automate the anomaly assessment, the quicklook imagery is scanned using a digital imaging camera system. The image which is generated is input to the "Anomaly Assessment System", which is resident on an image analysis system (Figure 1).

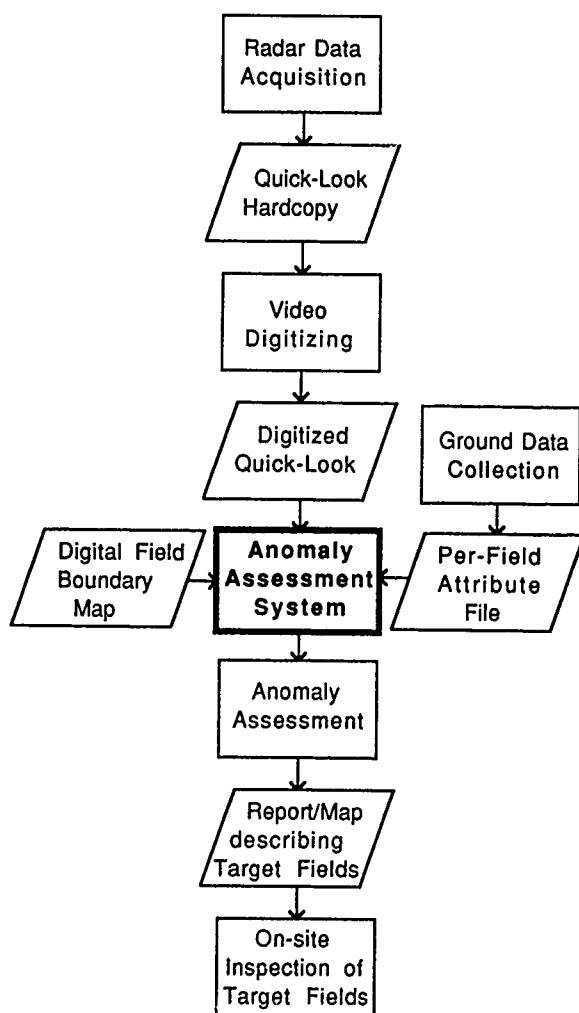


Figure 1: A flowchart of the Anomaly Assessment System

The second input to the Anomaly Assessment System is a digital field boundary map. Field and crop boundaries in the study area must be digitized and the file must be entered. The digitizing can be done using aerial photographs or a map. Alternatively, an existing digital file can be used. If necessary, the quicklook imagery when it is first received can be digitized to produce the field boundary file. The imagery can also be used to update any existing maps. This is important if a digital file from a previous year is to be used as the base map because crop boundaries do change from year to year. Once the map has been entered, the image must be registered to it.

The third and final input is ground data collected in the study area. An attribute file must be established containing information on crop types and their characteristics on a field-by-field basis.

Anomaly assessment consists of using masks to display on the colour monitor of the image analysis system only the fields with certain attributes, such as a specific crop type. This permits a more effective visual comparison of the fields and identification of the anomalies. In addition, summary maps and/or reports of statistics on a per-field basis can be generated and displayed. The quicklook imagery can be studied in detail to attempt to identify the reason for the anomaly. In addition, it should be possible to field check the study area within two or three days of the data acquisition.

Results

The procedure described above is operational using PCI image analysis software and the TYDAC SPANS Geographic Information System. It will be used as part of the data analysis strategy in an agricultural radar study that is currently taking place in Oxford County, Ontario. During the summer of 1988, CCRS acquired radar data over the test area on four separate occasions. On each occasion, data were obtained at different wavelengths, polarizations, depression angles and look directions. Digital field boundary maps for part of the study area have been generated and input to the system. In addition, attribute files for each field have been produced. With the above data and information sources, it is possible to do a retrospective anomaly assessment. It is hoped that during the summer of 1989 the system can be used in an operational mode to carry out anomaly assessment within a short period of time following data acquisition.

Conclusions

Anomaly assessment is an important element in the analysis of agricultural crops as they are recorded on radar imagery. The assessment permits greater understanding of the different factors influencing radar backscatter. The Anomaly Assessment System that is now in place will permit rapid checking of fields for anomalous tonal characteristics.

Acknowledgements

The research study reported in this paper was funded by a Centre of Excellence grant from the Province of Ontario to the Institute for Space and Terrestrial Science.

References

1. Cihlar, J., "On the relationship between agro-climatic region, crop parameters and SAR image tone in western Canada", *Can. J. Remote Sensing*, Vol. 12, No. 1, pp. 81-93, 1986.
2. Hirose, T., J. Cihlar and A. Mack, "Discrimination of agricultural crops in Canada with SAR plus VIR data; initial results", *Proc. 8th. Can. Symp. on Remote Sensing*, Montréal, pp. 387-403, 1983.
3. Ulaby, F.T. and P.P. Batlivala, "Optimum radar parameters for mapping soil moisture", *IEEE Trans. on Geosci. Electronics*, Vol. GE-14, 1976.
4. Ulaby, F.T. and T.F. Bush, "Monitoring wheat growth with radar", *Photogramm. Eng. and Remote Sensing*, Vol. 42, pp. 557-568, 1976.

MÉTHODE EFFICACE ET PEU COÛTEUSE DE CARTOGRAPHIE DES CULTURES PAR TÉLÉDÉTECTION SATELLITAIRE

Pierre Lafrance¹, Jean-Marie Dubois¹ et Manon Carignan²

1) Centre d'applications et de recherches en télédétection (CARTEL)
Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1
Téléphone: 819-821-7180 Télécopieur: 819-821-7238 Téléc: 05-836149

2) Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec
Service de conservation des ressources agricoles
200 A, chemin Sainte-Foy, 1^{er} étage, Québec, Québec, Canada G1R 4X6
Téléphone: 418-643-3029

Résumé

L'objectif de ce travail était de cartographier au 1:20 000, à l'aide de la télédétection satellitaire, les quatre cultures ou groupes de cultures suivantes: le maïs, les céréales, les fourrages et pâturages et les autres cultures. Il fallait de plus en calculer les superficies respectives au niveau régional. Le travail devait être fait dans un court laps de temps et à de faibles coûts. Aucune donnée de terrain prise simultanément au passage du satellite n'étaient disponibles. Le territoire traité est celui compris entre Pointe-au-Chêne et l'île du Grand Calumet en Outaouais, Québec. Des reproductions photographiques en fausses couleurs (TM5-rouge, TM4-vert, TM3-bleu) au 1:40 000 et au 1:80 000, d'une image Thematic Mapper du 11 août 1987, se sont avérées très efficaces pour planifier la visite sur le terrain et pour identifier les signatures visuelles des différentes cultures. Lors de la visite sur le terrain, au printemps 1988, il a été possible d'identifier, après la fonte de la neige, la plupart des cultures de l'été précédent. Les cas douteux pouvaient être éclaircis par des rencontres avec les cultivateurs. La date de l'image a une très grande influence sur le déroulement subséquent des travaux. Idéalement, elle doit être prise le plus près possible de la maturité des céréales sans que la récolte de celles-ci ne soit commencée. Sur le plan technique, il est intéressant de noter que les diapositives de l'image classifiée se superposent parfaitement aux cartes au 1:20 000 sans corrections géométriques. Enfin, la méthode que nous avons utilisée est très efficace et peu coûteuse et nous n'hésitons pas à la recommander pour des travaux d'inventaire.

Abstract

The object of the project consisted in the application of satellite remote sensing for mapping, at the scale of the 1:20 000, the four following crops or groups of crops: corn, cereals, hay and pastures and other crops. Furthermore, it was our aim to estimate, at a regional scale, the corresponding areas for each crop type. The work had to be done within a short period of time and at low cost. No ground data was acquired during satellite overflights. The area under study is located between Pointe-au-Chêne and île du Grand Calumet in the Outaouais region of Quebec. False colour photographic reproductions (TM5-red, TM4-green, TM3-blue), at the scale of the 1:40 000 and the 1:80 000, obtained from a Thematic Mapper image acquired on August 11, 1987, were very useful for planning field surveys and for identifying the visual signatures of the different crops. In the spring of 1988, after the snowmelt, it was possible to identify during a field survey most of the crops of the previous summer. Problem cases could be solved by meeting with the farmers. Image date had considerable impact on the subsequent steps to be undertaken. Ideally, it should be acquired closest to crop maturity, before harvest has begun. From a technical standpoint, it is interesting to note that the slides of the classified image can be perfectly superimposed on the 1:20 000 scale maps without applying geometrical corrections. Lastly, the method used is reliable and cost effective, and we do not hesitate to recommend the method for conducting inventory work.

Mots clés

Agriculture, cartographie, Thematic Mapper, classification

Key words

Agriculture, cartography, Thematic Mapper, classification

Introduction

Un important projet de recherche et de démonstration a été lancé par un consortium québécois dans le but d'évaluer les possibilités de valorisation agricole des boues de stations d'épuration des eaux usées municipales (Gosselin et St-Yves, 1988). Le Groupe-conseil Roche ltée s'est vu confié une partie du travail, dont l'évaluation des terres agricoles disponibles pour l'épandage. Le cas à l'étude est celui de la station d'épuration de la communauté régionale de l'Outaouais située à Gatineau, Québec.

Le Groupe-conseil Roche ltée a fait appel au Centre d'applications et de recherches en télédétection (CARTEL) de l'Université de Sherbrooke pour réaliser la cartographie des terres agricoles de la région comprise entre Pointe-au-Chêne et l'île du Grand

Calumet. Le mandat confié au CARTEL était donc de cartographier au 1:20 000, sur fond cadastral, les quatre cultures ou groupes de cultures suivants: le maïs, les céréales, les fourrages et pâturages et les autres cultures. De plus, leurs superficies respectives devaient être évaluées.

Compte tenu des délais très courts, des contraintes budgétaires et de la nécessité d'obtenir une cartographie la plus à jour possible, seule la télédétection satellitaire pouvait répondre à l'appel.

Méthode

Nous nous sommes inspirés d'une méthode mise au point au Centre canadien de télédétection par Bernier et Dupont (1986). Cette méthode consiste essentiellement en une classification supervisée d'une

image du capteur Thematic Mapper (TM) du satellite Landsat. Elle s'est avérée efficace dans les régions agricoles du sud du Québec. Les deux principales modifications que nous avons apportées à cette méthode sont l'omission de la bande spectrale TM1 et de l'étape de correction géométrique. Nous avons jugé pouvoir apporter ces modifications sans mettre en cause la qualité des résultats et en réduisant les coûts et les délais du travail.

Tous les traitements ont été effectués sur un système de traitement d'images ARIES III de l'entreprise Dipix Systems Ltd, implanté sur un ordinateur MicroVAX II de Digital Equipment Corporation.

Nous avons donc acquis les bandes spectrales 3, 4 et 5 (rouge 0,63-0,69 μm , proche-infrarouge 0,76-0,90 μm et moyen-infrarouge 1,55-1,75 μm) de l'image TM 16/28 du 11 août 1987. Elle avait subi la correction géométrique rudimentaire du système MOSAICS du Centre canadien de télédétection, qui corrige les distorsions dues à la rotation et à la courbure de la terre, à la vélocité du miroir et à l'angle de vue. Elle avait aussi subi une correction radiométrique élémentaire qui minimise le lignage et augmente le contraste.

Après avoir lu l'image complète, nous avons créé une sous-image couvrant le plus exclusivement possible le territoire à l'étude, ceci afin de ne traiter que la surface strictement nécessaire et ainsi réduire les coûts de calcul.

L'étude des histogrammes des trois bandes spectrales nous a permis de définir les limites des étirements linéaires de contraste. Notre but étant d'obtenir le plus de contraste possible dans la végétation, nous avons pu saturer les extrémités des courbes respectivement en noir et en blanc et ainsi disposer de plus d'amplitude pour la végétation. Le but de l'étirement linéaire de contraste était de produire des photographies fausses couleurs au 1:40 000 et au 1:80 000 des principales zones agricoles de la région afin de planifier et de mener les rencontres avec les agronomes et la visite sur le terrain. Pour ce faire nous avons associé la bande TM5 au faisceau rouge, la bande TM4 au faisceau vert et la bande TM3 au faisceau bleu. Sur de telles images composées fausses couleurs la végétation verte vivante apparaît en vert contrairement aux composées fausses couleurs traditionnelles où elle apparaît en rouge. Cette particularité s'est avérée profitable au moment de présenter les images aux agronomes et aux agriculteurs qui y voyaient une représentation plus naturelle du territoire.

L'inventaire ayant été commandé au CARTEL au printemps 1988, nous n'avons pu recueillir de données de terrain simultanément au passage du satellite. Il fallait donc, huit mois plus tard, recueillir le plus de données possible sur les cultures de la saison 1987. Nous avons donc rencontré les agronomes régionaux et locaux et visité les principales zones agricoles de la région. Il nous a été possible d'identifier sur le terrain, après la fonte de la neige, la plupart des cultures de l'été 1987. Les champs de maïs et de céréales laissent en effet des traces caractéristiques et les champs de fourrages ou pâturages sont pour la plupart encore dans cet état. Les cas douteux pouvaient être éclaircis par des rencontres avec les agriculteurs.

Nous avons pu ainsi identifier, sur l'image, les signatures visuelles des différentes cultures et déterminer des sites d'entraînement pour la classification supervisée.

Nous avons dû créer treize classes et sous-classes afin de couvrir les différents états des cultures et les autres types d'occupation du sol. Le tableau 1 présente ces classes et sous-classes. Nous avons ensuite procédé à une classification par maximum de vraisemblance (programme ML d'ARIES III).

Les différentes sous-classes de céréales, de fourrages et pâturages et d'autres cultures n'ayant été créées que pour les fins de la classification automatique, au moment de produire les diapositives de l'image classifiée nous avons regroupé ces sous-classes en une classe pour les céréales, une classe pour les fourrages et pâturages et une classe pour les autres cultures.

Les diapositives 35 mm couleurs ont été créées avec une photo-imprimante Quick Colour Recorder de l'entreprise Imapro Inc. Nous avons produit une diapositive pour chaque carte cadastrale au 1:20 000 afin d'obtenir un maximum de définition lors de la projection. Cette dernière s'est faite avec un projecteur à diapositives standard équipé

| Classes | Sous-classes |
|------------------------|--|
| Maïs | |
| Céréales | Céréales 1 Céréales 2 |
| Fourrages et pâturages | Fourrages et pâturages 1 Fourrages et pâturages 2 Fourrages et pâturages 3 Fourrages et pâturages 4 |
| Autres Cultures | Pomme de terre Crucifères Tabac |
| Forêt | |
| Zones bâties | |
| Eau | |

Tableau 1: Liste des classes et sous-classes

d'une lentille de type zoom et placé de façon à ce que le faisceau lumineux soit perpendiculaire à la carte fixée au mur sur un fond blanc. Le traçage a été fait à la main en interprétant les relations entre l'image classifiée et le fond cadastral. La présence des limites cadastrales permettait dans de nombreux cas de fixer plus précisément les limites des champs agricoles.

Parallèlement au travail de traçage, nous avons fait calculer automatiquement les surfaces des différentes classes. Un facteur de correction a dû être utilisé pour les céréales à cause d'une sur-évaluation due à des facteurs que nous verrons plus loin. La figure 1 illustre schématiquement la méthode que nous venons de décrire.

Résultats

Les reproductions photographiques au 1:40 000 et au 1:80 000, rehaussées par étirement linéaire de contraste, se sont avérées très efficaces pour planifier et mener les travaux de terrain et pour identifier les signatures visuelles des différentes cultures. Elles s'interprètent de la même façon et avec autant de précision que des photographies aériennes aux mêmes échelles et offrent beaucoup de subtilité dans les cultures. Les points de repère classiques, comme les voies de circulation, les limites cadastrales, les groupes de bâtiments, les cours d'eau, les talwegs et les boisés, sont facilement identifiables et facilitent les travaux de terrain.

L'absence de données de terrain prises simultanément au passage du satellite nous plaçait dans une position apparemment précaire. Heureusement, nous avons constaté qu'il est encore possible, au printemps suivant la saison à l'étude, d'identifier la plupart des cultures. Le maïs est reconnaissable par les nombreux débris grossiers qu'il laisse sur le champs. Les céréales sont reconnaissables par des rangées bien nettes de tiges coupées. Les champs de fourrages et pâturages sont identifiables pour leur part par leur tapis très dense. Les exploitations de tabac sont reconnaissables par les ensembles de bâtiments caractéristiques qu'on trouve près des champs.

Les cas douteux et les autres cultures pouvaient être identifiés par les agronomes régionaux ou locaux ou par les agriculteurs qui se sont montrés très réceptifs aux images satellitaires. Ils assimilaient très rapidement les notions d'échelle et de signature spectrale des champs cultivés. L'apparence naturelle de nos images composées fausses couleurs y était sûrement pour beaucoup, du moins lors du premier coup d'oeil.

L'étape de délimitation des sites d'entraînement s'est déroulée très facilement. Les images TM offre une résolution spatiale permettant de délimiter précisément la plupart des champs. L'étape de classification automatique a, pour sa part, donné lieu à un minimum de confusion entre les différentes classes à cartographier.

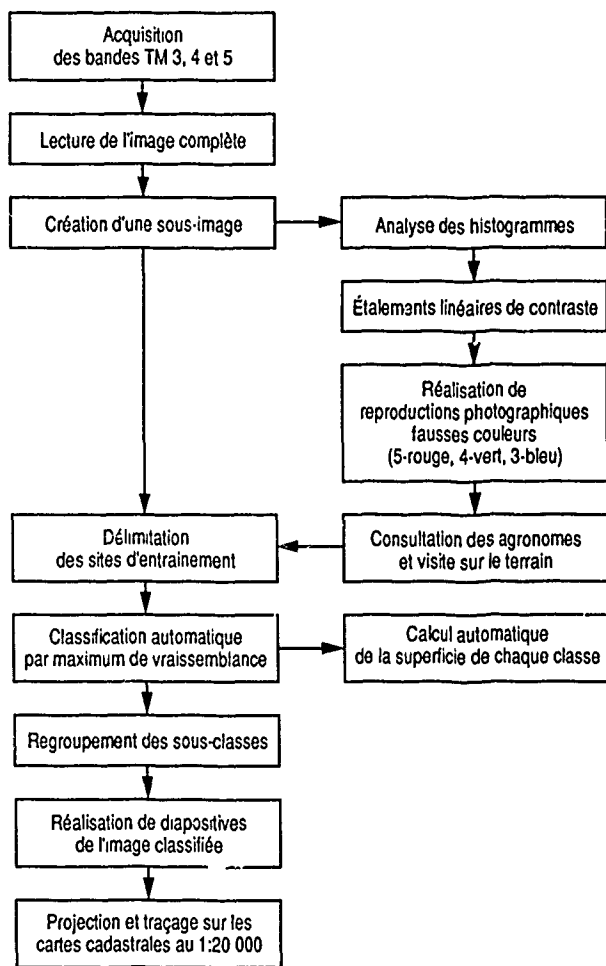


Figure 1: Schéma de la méthode

Au moment de choisir de ne pas faire de corrections géométrique, notre hypothèse était que l'image TM grossièrement corrigée serait assez juste pour se superposer presque parfaitement à une petite surface comme une carte cadastrale au 1:20 000. Cette hypothèse s'est avérée juste et seules quelques manipulations de projecteur et de lentille étaient nécessaires pour superposer les diapositives de l'image classifiée sur les cartes cadastrales. Les voies de circulation et les limites cadastrales étaient les principaux points de repère servant à guider la superposition. Encore là, la résolution spatiale de TM était bien adaptée à une cartographie au 1:20 000.

Discussion

Bernier et Dupont (1986) recommandent d'utiliser une image enregistrée entre la mi-juillet et la fin d'août, il nous semble cependant que l'image doit être choisie plus précisément en fonction de la région à l'étude et des caractéristiques climatiques de la saison en question. L'image choisie sera probablement située dans la majorité des cas entre la mi-juillet et la fin d'août mais, dans un cas particulier, une image du milieu de juillet ou de la fin d'août pourrait ne pas être satisfaisante du tout.

Le principal problème est l'identification des céréales. Trop tôt dans la saison, elles se confondent avec les fourrages et pâturages alors que trop tard elles sont récoltées et se confondent avec les surfaces non-végétalisées. À ce dernier moment, elles peuvent même se présenter dans plusieurs états différents: champs avec chaume, champ déchaumés, champs avec repousse de foin, etc.

L'image idéale se situe à la maturité des céréales mais avant que la récolte n'ait commencé. Ce moment dépend évidemment de la région et des conditions climatiques de la saison complète de culture. Il est facile de connaître les conditions climatiques d'une saison particulière et même la date approximative du début de la récolte en consultant les agronomes régionaux et locaux.

La saison 1987 a été particulièrement favorable en Outaouais et le 11 août plusieurs champs de céréales avaient été récoltés. Cette date était donc trop tardive dans ce cas particulier. Malheureusement, il n'existait pas d'autres images mieux adaptées à nos besoins. Les caractéristiques de passage du satellite Landsat et le climat canadien restreignent les probabilités d'obtenir plusieurs images sans nuages au cours d'une même saison de culture. On risque donc de se trouver assez souvent devant l'obligation d'utiliser des images pas tout à fait adaptées à l'application désirée.

Dans notre cas, il a été impossible de distinguer automatiquement les céréales de certains autres objets tels les jachères, les routes, les gravières et sablières, les dépotoirs et les résidus miniers. Au moment du report sur les cartes, il était facile d'identifier ces confusions par le contexte et de les corriger; par contre, au moment du calcul automatique des superficies, il a fallu évaluer la vraie superficie de la classe céréales en soustrayant un certain pourcentage associé aux objets avec lesquels il y avait confusion.

Nous jugeons opportun de rappeler l'importance de créer, au moment de la classification automatique, autant de classes pour chaque culture qu'il y a d'états dans lesquels cette culture se présente. Par exemple, le 11 août 1987, les champs de fourrages et pâturages pouvaient se présenter dans les états suivants: foin fraîchement coupé et ramassé, foin fraîchement coupé et couché sur le sol, foin en croissance, foin à maturité, etc. Il est donc impensable de demander à un algorithme de classification quel qu'il soit de classer une classe aussi large. Il est de loin préférable de faire autant de classes que d'états et de regrouper ces classes en leur assignant une même couleur au moment de la visualisation.

Conclusion

La méthode que nous avons utilisée pour ce travail est très efficace et nous n'hésitons pas à la recommander pour d'autres travaux du même genre. Les images du capteur Thematic Mapper, ayant une résolution spatiale de 30 m par 30 m, sont bien adaptées pour couvrir une région comme celle de la présente étude et offrent une précision des contours très satisfaisante. De plus, lorsqu'elles sont de bonne qualité et qu'en particulier on n'y trouve pas de voile d'humidité atmosphérique, elles contiennent beaucoup d'information sur le territoire agricole et les cultures.

Avant d'acquiescer une image, nous recommandons aux utilisateurs de s'informer auprès des agronomes régionaux et locaux et, si nécessaire, auprès des agriculteurs eux-mêmes, des conditions climatiques de l'année à l'étude et du moment approximatif de la récolte des céréales. Si une image prise juste avant ces récoltes était disponible, ceci éviterait la confusion entre céréales et sols nus.

La méthode de report par projection, bien que pénible pour les yeux du cartographe, est efficace et peu coûteuse. Il est à noter qu'il n'a pas été nécessaire de faire de correction géométrique à l'image TM pour qu'elle se superpose presque parfaitement aux cartes au 1:20 000. Ceci évite des manipulations longues et coûteuses en temps d'opérateur et de calcul.

Références

- Bernier, M. et O. Dupont, 1986. Réalisation de cartes de l'utilisation du sol en milieu agricole à partir des données de Landsat TM. Rapport méthodologique du projet conjoint TM/Agriculture (321044) présenté au ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec, Centre canadien de télédétection, Ottawa, Ontario, 51 p.
- Gosselin, Y. et A. St-Yves, 1988. La valorisation agricole des boues de stations d'épuration des eaux usées municipales. Municipalité, mars-avril 1988, p. 7-9

EVALUATION OF AGRICULTURAL MANAGEMENT APPLICATIONS IN THE RIO GRANDE VALLEY USING SATELLITE IMAGERY

Arthur J. Richardson*, Melba M. Crawford**,
James H. Everitt*, Lisa Dron**, and Rod Summy*

*USDA-ARS, Remote Sensing Research Unit
Weslaco, Texas 78596

**The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

The University of Texas at Austin and the USDA at Weslaco, Texas have begun a cooperative program, that has support from the local cotton industry, to evaluate agricultural applications of multispectral satellite imagery in the Rio Grande Valley (RGV). The primary application studied was that of monitoring defoliated cotton (Gossypium hirsutum L.) stalk destruction for pink bollworm (Pectinophora gossypiella [Saunders]) control. Satellite scenes acquired on July 17, October 2, and November 13, 1988, indicated that other useful agricultural management applications would be possible including mapping of saline soil and cotton root rot occurrence. Initial results indicate that 20 meter resolution multispectral satellite imagery is sufficient for on farm management decisions about individual fields.

KEYWORDS

Cotton, Pink BollWorm, Boll Weevil, Stalk Destruction

INTRODUCTION

Pink bollworm (Pectinophora gossypiella [Saunders]) and boll weevil (Anthonomus grandis Boheman) are the most destructive insect pests of cotton (Gossypium hirsutum L.) in the United States. Under the mild climate of south Texas cotton is a perennial such that cotton plants that are not destroyed after the late summer harvest provide a food source that supports a large population of bollweevils over the following winter (Summy et al., 1988). These weevils cause severe economic damage to the cotton crop during the next growing season. State legislation was passed in 1987 which requires all Texas cotton farmers to destroy the stalks in September after the harvest. Usually stalk destruction is done by plowing, hence the descriptive term plow-down to describe plant destruction for the cotton plow-down program.

Through proper management of the cotton plow-down program it is estimated that farmers in the Rio Grande Valley (RGV) could save as much as \$15 million in pesticide spraying costs and reduce pesticide usage as much as 400,000 gallons per year on the 300,000 acres of cotton generally grown in the RGV (Summy et al., 1988). Chemical runoff as well as direct and indirect negative environmental effects on humans and other crops would be greatly reduced as well.

Adequate monitoring is critical for effective management of the cotton plow-down program because, even though most farmers

comply with the law, the potential for large scale infestation by very small acreages of residual cotton is high. Currently, the USDA-ARS conducts high altitude flights and compares aerial color infrared (ACIR) photography from June-July with that from October and November to determine the location of cotton which has not been destroyed. A two year cooperative program between the University of Texas at Austin and the USDA-ARS at Weslaco, Texas, that has support by the local cotton industry, was initiated to investigate the use of high resolution multispectral satellite imagery for monitoring the cotton plow-down program in the RGV. In the course of this program additional agricultural management applications of potential benefit to farmers in the RGV have also been identified.

EXPERIMENTAL METHODS

A 6.4 by 4.4 km (4.00 by 2.75 mile) test site located between Elsa and Weslaco, Texas, (north latitude = 26.23°, west longitude = 97.96°) was selected to monitor the RGV cotton plow-down program using SPOT* multispectral satellite imagery acquired on July 17, October 2, and November 13, 1988. Aerial color infrared photographic imagery (ACIR) was also obtained for July 15, 1988. The cotton plow-down program and other possible agricultural applications visible in the satellite imagery were evaluated by comparison with the ACIR imagery.

RESULTS AND DISCUSSION

Green vegetation appears as red in the October 2, 1988, satellite false color image (Fig. 1). The shades of red change from light to dark corresponding to changes of vegetation amounts from sparse to dense. The satellite image is annotated by a quarter mile row and column grid system. Most vegetation shown in this October scene is either sugarcane, such as the field located at rows 5 and 6 in columns 6 and 7, or vegetables, such as the corn field located at rows 7 and 8 in columns 4 and 5. All cotton should have been harvested and plowed-down before September 15, 1988. However, there were two fields of undestroyed cotton remaining after the September 15, 1988, cotton plow-down deadline, that appear in the SPOT satellite imagery at rows 9 and 11 in column 1. There is a citrus orchard located between these two cotton fields.

Field visits in October showed that the cotton in these two fields (Fig. 1) had been defoliated and then harvested but had not been shredded and plowed down. Thus, the cotton left in these two fields refoliated producing the red signature characteristic of green vegetation and could harbor boll weevils over the 1988 winter period if not eventually plowed-down. These boll weevils could then become a problem for the 1989 summer cotton season. Most cotton fields in this area were properly plowed down, such as the field located at row 14 in columns 8 and 9, and is now void of cotton regrowth.

In actual operations satellite images before and after the September 15 cotton plow-down program deadline would be compared to locate cotton that has not been destroyed. The University of Texas at Austin, in cooperation with the USDA-ARS, Remote Sensing Unit, at Weslaco, is developing computer aided image processing techniques to make these comparisons automatically. When developed, these image processing techniques could become part of a satellite monitoring system that could save agriculture in the Rio Grande Valley millions of dollars a year in chemical costs for control of the pink bollworm.

The 20 meter resolution of the SPOT satellite imagery is also capable of mapping areas of cotton root rot (Nixon et al., 1975), a fungal disease (*Phymatutrichum omnivorum*) of the soil, and salinity (Everitt et al., 1988) in cotton and sorghum crops. Root rot and salinity cannot be monitored using October satellite images because these problems can be mapped only during the growing season when there is vegetation growing. However, one example of soil salinity effect can be seen in the light colored soil in the center of the field that contained cotton at row 14 in columns 8 and 9. Imagery acquired during the growing season, July 17, showed that the

cotton crop in this area was very sparse because of the soil salinity effects. Root rot also causes a decrease in plant cover similar to soil salinity effects.

Irrigated fields can be detected, such as the fields at row 2 and columns 7, 9, and 10 (Fig. 1) because wet fields appear as dark in the false color satellite imagery. The ability to detect excess soil moisture could be useful during periods following heavy rainfall or to detect and manage fields that are not draining properly.

Recent problems in the Rio Grande Valley with citrus black fly (*Coccus hesperidurum* L.) infestations can also be mapped using the satellite imagery. The dark vertical strip, located at row 8 and column 11 in the October satellite image (Fig. 1), is a citrus orchard just affected by citrus black fly. The black fly produces a black sooty mold fungus (*Capnodium citri* Berk. and Desm.) on the surfaces of citrus leaves resulting in the dark reflectance signature shown in this satellite image (Gausman, 1971). If untreated, the sooty mold reduces the productivity and economic value of a citrus orchard. Biological methods that utilize blackfly parasites is the method generally used to control black fly. A earlier July satellite and ACIR image indicated that the field was still normal in appearance.

Two acres of Kenaf (*Hibiscus cannabinus*), located at the north USDA-ARS research farm can be seen at row 9 in columns 10 and 11 (Fig. 1). The Kenaf is the small red square in the center of the farm that is largely bare of experimental crops at this time of year. Kenaf is being considered as a crop for the RGV that will be economically important for the production of paper. Thus, monitoring Kenaf could be a future agricultural management requirement for the RGV.

CONCLUSIONS

The cooperative program between the University of Texas at Austin and the USDA-ARS at Weslaco, Texas shows promise of bringing high technology aerospace techniques to the management of agriculture for the farmer. This technology should allow monitoring of many different agricultural management problems in the Rio Grande Valley of south Texas. The advent of 20 meter multispectral satellite imagery and the availability of inexpensive image processing and geographic information systems as well as inexpensive yet powerful microcomputer systems means that the remote sensing satellite technology developed over the last twenty years could become available for on farm as well as on ranch implementation.

ACKNOWLEDGEMENTS

Thanks are extended to Rene Davis for obtaining complete aerial photographic coverage of the Rio Grande Valley for July 15, 1988, of which a portion was utilized for this report. The authors are also indebted to Rene Davis for his expert

*Trade names are included in this manuscript for the benefit of the reader and do not imply a preference for the products listed by the U. S. Department of Agriculture.

knowledge of agricultural management problems in the valley and how they relate to aerial photographic information and coverage. We acknowledge the image processing expertise of the following University of Texas student researchers: Sanghoon Lee, Paula Frazier, and Tom Kleeman. The authors thank Wayne Swanson for his help with the computer graphics.

REFERENCES

- Everitt, J. H., D. E. Escobar, A. H. Gerbermann, and M. A. Alaniz. 1988. Detecting saline soils with video imagery. *Photogram. Eng. and Remote Sens.* 54(9):1283-1287.
- Gausman, H. W. 1971. Photographic remote sensing of "sick" citrus trees. *Proc. Earth Res. Survey Systems.* II:15-30.
- Nixon, P. R., S. D. Lyda, M. D. Heilman, and R. L. Bowen. 1975. Incidence and control of cotton root rot observed with color infrared photography. *Tex. Agric. Exp. Sta. Misc. Publ.* 1241. 4p.
- Summy, K. R., Hart, W. G., Davis, M. R., Cate, J. R., Norman, J. W., Wofford, C. W., Heilman, M. D., and Namken, L. N. 1988. Regionwide management of boll weevil in southern Texas. Beltwide Cotton Production Research Conference.

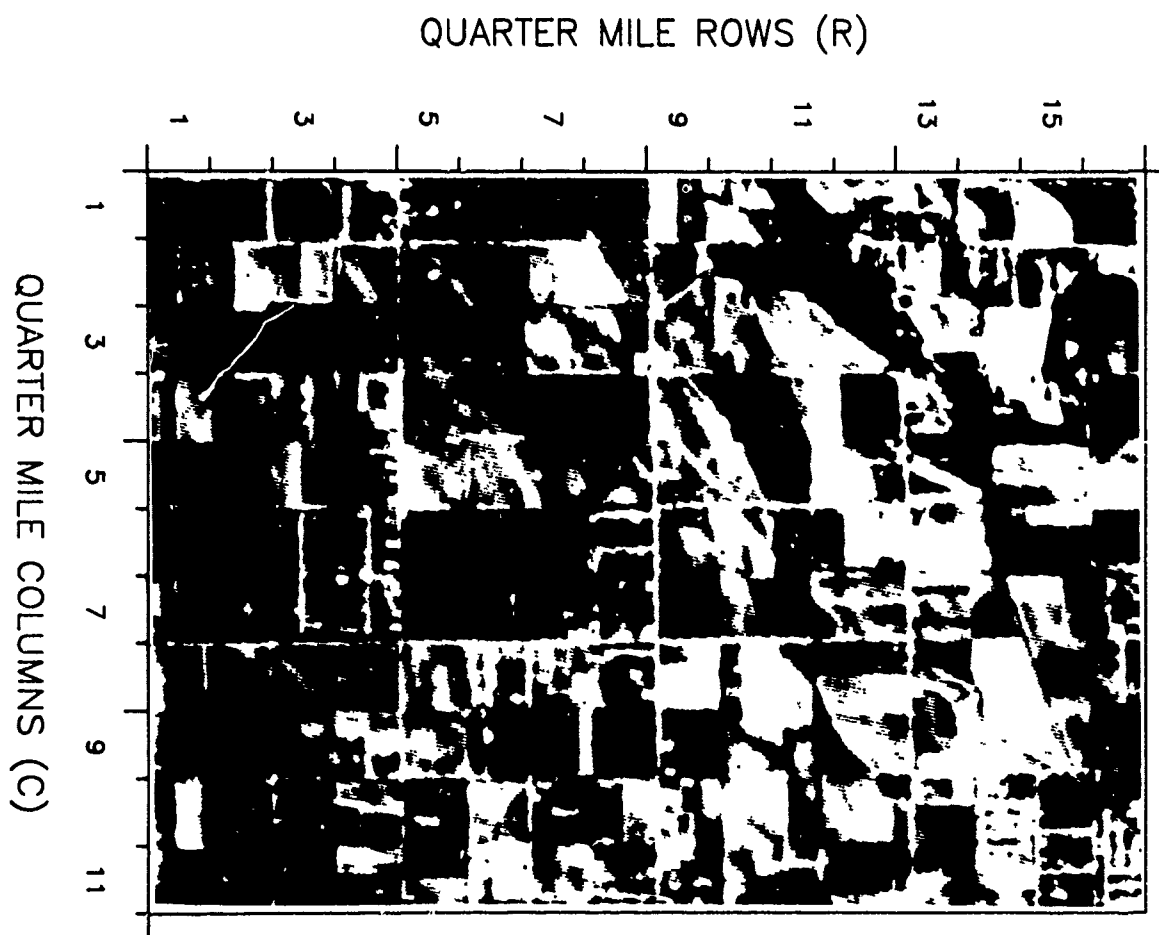


Figure 1. Multispectral scanner SPOT imagery for cotton plow-down study area in the Rio Grande Valley of south Texas acquired on October 2, 1988. Two fields of undestroyed cotton remained after the September 15, 1988, plow-down deadline at rows 9 through 11 in column 1.

COOPERATIVE SYSTEMS FOR PERCEPTUAL TASKS IN A REMOTE SENSING ENVIRONMENT

Alan K. Mackworth

Fellow, Canadian Institute for Advanced Research
Department of Computer Science
University of British Columbia
Vancouver, B.C. V6T 1W5

Abstract

To design and implement knowledge-based systems for perceptual tasks, such as interpreting remotely-sensed data, we must first evaluate the appropriateness of current expert system methodology for these tasks. That evaluation leads to four conclusions which form the basis for the theoretical and practical work described in this paper. The first conclusion is that we should build 'cooperative systems' that advise and cooperate with a human interpreter rather than 'expert systems' that replace her. The second conclusion is that cooperative systems should place the user and the system in symmetrical roles where each can query the other for facts, rules, explanations and interpretations. The third conclusion is that most current expert system technology is *ad hoc*. Formal methods based on logic lead to more powerful, and better understood systems that are just as efficient when implemented using modern Prolog technology. The fourth conclusion is that, although the first three conclusions can be, arguably, accepted for high-level rule-based symbol-manipulation tasks, there are difficulties in accepting them for perceptual tasks that rely on visual expertise. In the rest of the paper work on overcoming those difficulties in the remote sensing environment is described. In particular, the issues of representing and reasoning about image formation, map-based constraints, shape descriptions and the semantics of depiction are discussed with references to theories and prototype systems that address them.

Keywords: Cooperative systems, expert systems, map interpretation, remote sensing, perceptual tasks, shape description, depiction.

1. Introduction

The aim of this paper is to present a research program underpinning the design and implementation of cooperative systems for perceptual tasks in a remote sensing environment. The need for the program is motivated by an analysis of current expert system technology and the requirements of perceptual tasks.

2. Expert Systems

Expert system technology has extended the range of problems amenable to computer-based solutions. The classic view of an expert system is that it is a computer program that satisfies the following requirements (Walker *et al.*, 1987). It must solve, or help to solve, an important problem that would otherwise require human expertise and judgement. It must integrate new knowledge incrementally into the knowledge base. It must help the designer and user elicit, organize, display and transfer knowledge. It must provide explanations of its advice. It must reason with inexact and exact knowledge. Finally, it must support a readable and natural user interface. To satisfy these requirements an expert system must have a knowledge base of facts and rules, and a rule engine for deriving new facts and establishing goals. In addition it may have an explanation generator, methods for acquiring and encoding new knowledge and a dialogue handler for the user interface.

The standard task classification for expert systems (Stefik *et al.*, 1982), breaks the applications into the following six generic task domains: interpretation, diagnosis, monitoring, prediction, planning and design. Problems in these domains have in common the characteristic that their space of possible solutions is very large, ruling out a generate-and-test exhaustive enumeration algorithm. In addition the tasks may require tentative

reasoning with incomplete knowledge on time-varying, noisy or incomplete data.

On the face of it then, current expert system technology, as embodied in rule-based shells, is ideal for interpretation tasks in a remote sensing environment such as, say, updating a forest cover map based on satellite imagery, a digital elevation model and an earlier forest cover map. But, in fact, that is far from the case; we shall examine why this is so and what can be done about it.

3. Cooperative Systems

There are essentially two reasons why current expert system technology is inappropriate for most perceptual tasks. First, it doesn't work very well. Second, it takes us in the wrong direction. It doesn't work for a variety of reasons, but basically because it fails to build internal models of the process it is trying to understand.

Attempts to build diagnosis systems that reason about the structure and function of the device to be diagnosed are leading to better performance. This is known as 'diagnosis from first principles' or 'model-based reasoning'. We must do the same for interpretation tasks.

When I say it takes us in the wrong direction, I mean simply that trying to build a program to replace a human expert represents an attempt to de-skill the task. It is not likely to lead to strong support from the expert; moreover, if the program cannot interact in terms of models the expert understands the prospect of effective knowledge transfer is minimal.

This rationale lies behind arguments for designing and building prototype 'cooperative systems' that advise and cooperate with an expert, or a novice, interpreter. Such systems are 'cooperative' in another sense as well. We envisage a cooperative system having a variety of knowledge sources, *including the user*, and allowing them to cooperate to arrive at a mutually consistent interpretation.

For example, given that standard maximum likelihood methods can produce partially correct classification results based on spectral signatures (subject to several restrictive assumptions), the user could sketch a map on the image, allowing a sketch map interpretation program to interpret the map, supplying spatial constraints and context sensitivity to a spectral segmentation knowledge source. Such a cooperative system has been realized (Glicksman, 1983; Havens and Mackworth, 1983).

For such systems to succeed, the user, who is an active participant in a cooperative system, must come

to trust the other components. This will only occur if the other components are seen to be *transparent* and *reliable*. Perhaps the only way to ensure this is to place the user and each component in symmetrical roles where each can query the other for facts, rules, explanations, interpretations and justifications. This approach to interfaces for knowledge-based systems has been called the QUARFE interface since it is based on a protocol that allows Questions, Answers, Rules, Facts and Explanations (van Emden, 1988).

Much current expert system technology is *ad hoc*. The programs are large, complex, opaque, and unreliable. When an answer is computed its relation to the input is unclear and when new rules are added they interact with old rules in unexpected ways. These problems lead to major difficulties in scaling up from small projects to large ones. By implementing the rule interpreter in Prolog as an extension of the standard Prolog interpreter (Sterling and Shapiro, 1986) one can add facilities that implement the QUARFE interface. This also allows one to know that a conclusion drawn by a component is a logical consequence of its facts and rules. This enables the user to trust the component because it is transparent and reliable. Moreover, one can implement, in the rule interpreter, schemes for representing inexact knowledge based on a formal Bayesian theory of probability without paying the price of losing the clear semantics of first order logic (Poole, 1989).

4. Remote Sensing Tasks Requirements

Although the conclusions of the previous section regarding the need for, and structure of, cooperative systems may be, arguably, accepted for high-level symbol manipulation tasks there are major hurdles on the path of implementing them for perceptual tasks that rely on visual expertise. In this section we mention some of those difficulties and describe some work on overcoming them.

A cooperative system whether it be for diagnosis or interpretation must represent, and reason about, the underlying physical reality it is dealing with by constructing an adequate model of that reality. For perceptual interpretation tasks the image formation process must be thoroughly modelled. We cannot make simple assumptions such as assuming that image irradiance is a function solely of a scalar surface albedo which uniquely characterizes the ground cover. Models of the optics of image formation including distributed illumination sources, scattering and haze, surface elevation, slope and aspect, non-Lambertian surface reflectance and so forth as, for example, in (Woodham *et al.*, 1985) and

(Woodham and Gray, 1987) are essential.

These models capture the optical constraints of the task, and they are necessary but they are not sufficient. There are often non-optical, physical and perhaps even cultural, constraints that must be modelled. The user and the various components of the cooperative system must be able to communicate questions, answers, rules, facts and explanations about such constraints easily. Although, a component can model these constraints *internally* using sentences in predicate calculus, at the user interface they must be represented visually using the graphical formatting conventions that have developed into the language of maps.

In our Mapsee project we have designed, implemented and tested a series of computer programs, Mapsee-1, Mapsee-2 and Mapsee-3, for interpreting sketch maps of geographic regions. Mapsee-2 and Mapsee-3 use schemas to represent and reason with a variety of constraints (e.g. 'roads must be on land', 'a shoreline bounds a land region on one side and a water region on the other', . . .) to achieve a consistent interpretation of the map. See (Mulder *et al.*, 1988) for an overview of the Mapsee project. As mentioned in Section 3, Mapsee has been used to provide a visual interface to an interpreter, allowing him or her to sketch features over aerial images of small towns (Glicksman, 1983). The Mapsee interpretation of the sketch is then used to provide tight contextual constraints to a traditional maximum likelihood classifier. This illustrates the cooperative principle: the user and the system components each contribute their expertise to arrive at a shared goal.

In general then, a cooperative system for remote sensing tasks must accept images in various forms as 'facts' from the user. In order to be useful these images must be referred to a canonical coordinate system and interpreted into a common framework. But we should not require the user to provide a map registered to an image or to carry out manual registration through the use of ground control points. Sensor-based raster imagery must be automatically registered to map-based vector data. Both the raster imagery and the vector imagery may be provided at many different scales. Good descriptions and multiscale matching techniques are necessary for registration. Recently we have developed techniques that smooth contours in a path-based coordinate system. These techniques have certain properties that are essential for matching purposes (Mackworth and Mokhtarian, 1988). We have also developed a system that exploits this representation to achieve automated registration of Landsat MSS

data to a map database (Mokhtarian and Mackworth, 1986).

The user interacts with a perceptual cooperative system visually – by drawing and reading sketch maps, for example, rather than typing sentences. On the other hand, we concluded in Section 3 that such a system should use a formal logic as its internal representation language. (Or, at least, it can be characterized as so doing – it may not be implemented that way.) This leaves a chasm to be bridged: we need a logical theory of the semantics of maps. We describe such a theory for diagrams in general in (Reiter and Mackworth, 1989). This logical framework requires the implementer to write a set of sentences in first order predicate calculus to describe the image, another set to describe the scene and a third set to describe the image-scene depiction mapping. These sentences can include general facts about a class of images (the set of all maps, say) – what image objects can appear in them, how they relate and so on – and also facts about a particular image. An interpretation of a diagram is defined to be a logical model of the complete set of sentences. An implementation of this theory now exists in prototype form for a restricted class of maps.

5. Conclusion

In summary, knowledge-based systems for perceptual tasks should be designed as cooperative systems not expert systems. Cooperative systems should be seen as including the user. The components of such a system interact in symmetrical ways using the QUARFE protocol. Internally, the components should be specified and, perhaps, implemented using first order predicate calculus.

An analysis of the obstacles preventing the implementation of such systems for perceptual tasks in a remote sensing environment was presented. This analysis established four research goals. The first is to build adequate models of image formation. The second is to represent and reason about map-based constraints. The third is to derive multi-scale shape descriptions and matching algorithms. The fourth is to establish a theory of depiction that applies to maps and other diagrams. The resulting theories and prototype systems demonstrate that the goal of building cooperative systems for perceptual tasks is achievable.

6. Acknowledgements

The financial support of the Natural Sciences and Engineering Research Council of Canada and the Cana-

dian Institute for Advanced Research is gratefully acknowledged. I am also grateful to the faculty, staff and students of the Laboratory for Computational Vision for ongoing dialogue and support.

References

1. Glicksman, J., "Using multiple information sources in a computational vision system", *Proc. IJCAI-83*, Karlsruhe, W. Germany, pp.1078-1080, 1983.
2. Mackworth, A.K. and Mokhtarian, F. "The renormalized curvature scale space and the evolution properties of planar curves", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Ann Arbor, MI, pp.318-326, 1988.
3. Mulder, J.A., Mackworth, A.K., and Havens W.S., "Knowledge structuring and constraint satisfaction: the Mapsee approach", UBC Dept. of Computer Science, TR87-21, 1987.
4. Poole, D. "Computer Science 532 course notes", Dept. of Computer Science, University of British Columbia, 1989.
5. Reiter, R. and Mackworth, A.K. "A logical framework for depiction and image interpretation", UBC Department of Computer Science, TR88-17, to appear in *Artificial Intelligence*, 1989.
6. Stefik, M. *et al.*, "The organization of expert systems, a tutorial" , *Artificial Intelligence* 18, pp.135-173, 1982.
7. Sterling, L. and Shapiro, E. *The Art of Prolog: Advanced Programming Techniques*, MIT Press, Cambridge, MA, 1986.
8. van Emden, M.H., "Project 4.5", University of Victoria, Victoria, B.C., 1988.
9. Walker, A., McCord, M., Sowa, J.F. and Wilson, W.G. *Knowledge Systems and Prolog*, Addison-Wesley, Reading, MA, 1987.
10. Woodham, R.J., Catanzariti, E. and Mackworth, A.K., "Analysis by Synthesis in Computational Vision with Application to Remote Sensing", *Computational Intelligence*, 1(2), pp.71-79, 1985.
11. Woodham, R.J. and Gray, M.H., "Analytic method for radiometric correction of satellite multispectral scanner data", *Proc. IEEE-GE*, vol.25, pp.258-271, 1987.

EXPERT SYSTEMS AND ENVIRONMENTAL CHANGE - INFORMATION EXTRACTION
FROM REMOTE SENSING DATA

David G. Goodenough¹, Michael A. Robson², and Ko B. Fung¹

¹Canada Centre for Remote Sensing
Energy, Mines and Resources Canada
Ottawa, Ontario Canada

²INTERA Technologies Ltd.
Ottawa, Ontario, Canada

ABSTRACT

Techniques used in remote sensing applications for resource mapping can be used over long periods of time to provide information on dynamic, environmental processes of geographical areas of interest. Expert systems will be useful in extracting environmental predictions from complex sources of data and in automating the processes of data integration, object representation, image interpretation, interpreting the results in terms of environmental parameters, and determining the changes in the parameters over time. Additional predictions which can be extracted from the same knowledge base are: when and where further measurements are required, which measurements are most essential, and predictions of which changes are expected in specific areas given the past history and rates of change.

Representations of data within the expert systems will include multi-temporal remote sensing from multiple sensors and platforms, GIS data, digital terrain models, field measurements, and environmental parameters. Knowledge of information extraction will include data integration, image analysis, physical and biophysical interpretation of ground measurements, and extraction and interpretation of environmental parameters. This paper examines methods for data and knowledge representation for environmental change as detected with remote sensing data and describes a design for a hierarchical expert system with distributed knowledge and expertise. Given the complex nature of the environmental problem of terrestrial warming, expert systems will be required to blend the diverse data sources, models, and to make predictions.

1.0 INTRODUCTION

Remote sensing is being used for renewable and non-renewable resource mapping, monitoring, and exploration. The techniques used in these applications of remote sensing can provide over long periods of time information on dynamic, environmental changes in geographical areas of interest. Information from Geographic Information Systems (GIS) and biological ground measurements, in combination with remote sensing data and models, can provide information over large areas about environmental changes from which future predictions about the environment can be made. Modeling of the earth's climate under the assumption of a doubling of CO₂ leads to predictions of large temperature increases and major shifts in precipitation patterns (Parry *et al*, 1988) resulting in major changes

of vegetative cover, especially in sensitive biomes.

A monitoring of the environment of a substantial area, such as a continent, will require the integration of data, information, knowledge, and expertise from many sources. Expert systems will be useful in automating the processes of data integration, object representation, image interpretation, interpreting the results in terms of environmental parameters, and in determining the changes in these parameters over time. Expert systems can also assist in extracting environmental predictions from complex sources of data of varying accuracies. Integrated expert systems can also serve as a decision support tool to decide when and where further measurements and data are required, which measurements are most valuable, and to even suggest options for actions to minimize environmental damage.

Representations of data within the expert systems will include multi-temporal remote sensing from multiple sensors and platforms, GIS data, digital terrain models, field measurements, and environmental parameters. Knowledge of information extraction will include data integration, image analysis, physical and biophysical interpretation of ground measurements, and extraction and interpretation of environmental parameters. No single expert exists who covers all of these areas. Therefore, knowledge within the expert system will be associated with different levels of expertise and scope and will come from multiple human experts.

We propose an hierarchy of experts which we call the Environmental Monitoring System (EMSys). Initially, each level of knowledge (or expert) in EMSys will have the ability to draw upon the knowledge from lower levels. Any lower level expert may be required to provide information to any number of higher level experts. Experts at the lowest level will be required to perform generic operations with the actual data (remote sensing, GIS, ground measurements). The next higher level will interpret these data and attempt to extract environmental parameters of interest. The next higher level will make environmental predictions based on models, interpreted data and historical information. Several intermediate levels will be required to organize experts with different scopes of knowledge. For example, experts which interact purely with remote sensing data will be separated from experts which interact purely with ground measurements, although they may be at the same level of expertise.

The low level of experts will perform multi-sensor integration by combining remote sensing data from different sensors which have different spectral, spatial, and temporal characteristics. This level will also integrate remote sensing with GIS and digital elevation data, perform image analysis on the combined data set, and perform preliminary analysis on the ground measurements.

The higher level of experts will require the most extensive use of uncertainty in the system. The certainty computations which have the most value will be: the confidence that a computed change in the environmental parameters is significant, the probability that a detected change is an effect of climatic change, and the contribution that any given measurement variation has had in terms of estimating uncertainties for environmental predictions. The environmental history of a given geographical site will also be treated at the higher level of experts. Given the complex nature of the environmental problem of terrestrial warming, expert systems will be required to blend the diverse data sources, models, and to make predictions.

2.0 SOURCES OF ENVIRONMENTAL DATA

Environmental monitoring with remote sensing can be carried out by recording the current status of the environment, recording the status at a future time, and identifying environmental changes. For some environmental changes, the interval to detect important changes may be unacceptably long when compared with the time to implement corrective actions. An alternative approach, therefore, may begin with the recording of the current status of the environment, followed by the prediction of environmental changes likely to take place. Successive remote sensing observations over time (snap shot records) can be used to confirm these predicted changes. Ground measurements, based upon a stratified, random sampling will still be essential in order to detect quickly subtle biophysical changes not yet visible in the spectra of the canopy. The information derived from remote sensing and ground measurements will be stored in geographic information systems.

2.1 Remote Sensing Sensors

Remote sensing sensors aboard aircraft and satellites can provide snap shots of the earth in the form of images taken in the optical, infrared, and microwave spectral regions. It is unlikely that any single sensor or platform will be sufficient to provide all of the environmental information required. The environmentalist will have to deal with thousands of images from a wide variety of sensors. The detected objects will be compared with previous information. This earlier information will likely have been derived from aerial photographs which had been used to make maps and to provide resource information. Since the remote sensing observations may span a wavelength range from 400 nm to 22 cm, different properties of objects, and even the kinds of objects detected, will change with sensor. For example, high resolution multispectral imagers may provide accurate forest species information, while microwave radars would indicate surface roughness, crown density, and moisture conditions.

Unfortunately, no data source is free from errors. The suppliers of remote sensing data, such as the Canada Centre for Remote Sensing (CCRS), expend considerable resources in keeping the calibrations of their sensors as accurate as possible. However, sensors will have

changes in radiometric sensitivities over time. There will also be the occasional missing line or portion of a line. Users can order Canadian geocoded products, for example, where geometric corrections have been made using ground control derived from 1:50,000 NTS maps, which themselves have variable accuracies over time.

In order to use these data effectively, one must include many other factors in the information extraction process. These include the viewing geometry, the illumination sources, the atmospheric conditions under which the data were collected, and the effects of topographic relief. Viewing geometry and the local terrain relief cause occlusions, shadows, layover in microwave imagery, positional shifts of pixel locations, and apparent compression or elongation of objects, depending upon the sensor being used.

For optical sensors it is essential to correct for atmospheric effects as precisely as possible in order to detect the most subtle changes. Often this means the use of atmospheric models and the integration of observations from meteorological satellites and ground stations. Environmental monitoring is particularly important for the most sensitive biomes, such as the boreal and alpine forests in Canada. Observations of the environment in mountainous terrain makes further demands on the atmospheric models in terms of opacity as a function of altitude.

From this brief list of factors, one can easily see that the environmental scientist is going to be overwhelmed by the large amount of information and complexities of the various sensors and platforms. The search space of possible interpretation paths and corrections is large. Expert systems are useful in simplifying the user's interaction with a complex search space and in making that search more efficient. For a number of these problems, we have recorded knowledge of the sensors, their corrections, and analysis procedures. Procedures have been written to deal with some of the integration problems and will be described in more detail in section 3.1.

2.2 Geographic Information Systems

Geographic Information Systems (GIS) are used to organize, store, analyze, and disseminate geographically-referenced environmental data. Information which can often be found in geographic information systems include current and historical distributions of forest cover types, other vegetation, wildlife habitat, soils, topographic relief, environmental disturbances, and, sometimes, climatic records. A major task facing designers of a national environmental monitoring system is the integration of information from a wide variety of national and international GIS. There are many obstacles to standardization of GIS. The large investments in hardware, software, and data acquisition cause an almost stifling inertia. Many GIS use different hardware, exchange protocols and media, and different representation formats for output and displays. Even for GIS from the same vendor, users will have established different data structures with different meanings and attributes for such features as forest, roads, rivers, creeks, agriculture, etc.

In most geographic information systems, raw data, usually in the form of survey notes, aerial photographs, and sometimes remote sensing data, are transformed cartographically to a vector based representation from which one can produce paper maps with legends. During this transformation, the data are

edited according to operational rules, and symbolized according to predefined standards. Sometimes data may be omitted due to certain practices, such as minimum mapping areas, current importance of such features, etc. Therefore, the archived products are densely packed symbols, graphics, and database showing the relevant information for the operational agency. This cartographic process is a mixture of art and science. The product is efficient for human utilization, but makes multisource integration more difficult. Current research is directed to circumventing this problem and some results are discussed in Section 3.1.

2.3 Biophysical Ground Measurements

Given the complexities of the interpretations expected of the remote sensing data, it is essential that there are frequent biophysical ground measurements. These measurements provide a reference for parameters that can be derived from remote sensing and a source of important biological parameters which can not be remotely sensed. These measurements would include atmospheric properties, soil conditions, gas exchanges, plant cell analyses, ages, biomass distribution, and so forth. There would also be ground measurements of solar illumination and spectral reflectances and microwave scattering of surface objects. The measurements would need to be collected in a statistically valid fashion, presumably on a stratified grid with random selection of measurement sites within grid cells.

The biophysical ground measurements would be incorporated into models and into the primary GIS. Within the GIS it would be necessary to scale up these measurements to cover areas sufficiently large that meaningful comparisons with remote sensing observations could be made. For many of these measurements, there is a great deal of uncertainty as to how to do this scaling. EMSys would need to support interpretations of integrated data for which there would be large uncertainties.

Historical geographic information in the GIS can serve to constrain broader interpretations and to enable unbiased estimates of accuracies. The biophysical measurements will provide the details necessary for precise evaluation of current interpretations and predictions.

3.0 ENVIRONMENTAL INFORMATION INTEGRATION

3.1 Information Integration

An environmental monitoring system should keep track of objects over time. Ambiguities arise when one attempts to define precisely such objects as, for example, forest stands. That which is a forest stand of some grouping of species will likely evolve under the stress of climatic change to a different species mix or even to a non-forest state. Polygonal representations are poor for tracking such temporal changes. For this purpose the most effective tessellation of the earth is a grid. We expect, therefore, that EMSys will utilize a variety of grids. Polygonal representations will be generated as required to depict environmental objects for specific times.

Objects have attributes which vary over time. These attributes include spatial and spectral electromagnetic responses, size, biomass, orientation, location, and so forth. If we consider objects as being represented by an n -dimensional vector, then remote sensing measurements represent an incomplete sample of these attributes for a specific time. Information

integration is intended to make the description of object attributes more complete. By doing so, one hopes to be able to more precisely recognize and characterize environmental objects, and to be able to better detect subtle changes. The integrated data would be transformed into information. Some objects would be well identified at coarser resolutions and fewer spectral channels. Thus, the integrated data would use iterative, interpretation techniques. An important need in planning the interpretation strategy is a means to assess the information value of a particular sensor or GIS in a quantitative manner. The value of the source will depend upon the queries being answered by EMSys.

Integrating data from different sources of measurements is not simple. A common technique for integrating data from different sources is to rectify the data to a common map projection and to overlay them in a data pile stack. This technique suffers the pitfalls of positional shifts, layover problems, and variable ground resolution in the case of obliquely looking sensors and side looking radar images. In order to remove these kinds of distortions, digital terrain models (DTM) are used to provide a correction mechanism for most of the distortions. This complex topic is the subject of another paper at this conference (Goodenough et al, 1989). The accuracy of the DTM becomes one of the limiting factors in the data integration. Also, there is a substantial computational cost associated with geometrically correcting all data sources. In EMSys this computational cost will only be incurred when the value of the source for the desired information warrants it.

The information extraction process will be performed on the appropriate data stack(s). Temporal analysis using a stack pile of images from different sensors at different times is a powerful tool in tracing biological cycles and determining future trends. Effects due to sun angle differences, viewing geometry, and temporal changes of atmospheric scattering can adversely affect the accuracies of the interpretations. At CCRS these effects are minimized using atmospheric and sun angle correction software.

The information extraction process is a complex procedure which requires the knowledge of experts. Unlike simpler expert systems, EMSys will need to be able to evolve as human experts become more expert. The model of a static human expert, so often cited in AI literature, is not applicable to the environmental monitoring problem. EMSys will itself need to incorporate a learning ability. At the present time this is beyond our capability.

Current research being performed by the authors looks into goal-directed information integration. We postulate that instead of classifying a large stack of images, it is possible to identify objects from a few images and then verify their accuracies with other imagery or geographic information. The system will include a knowledge base containing plans and accuracies of classified images. When a goal is entered into the system, EMSys will first search through the plans and identify the most cost effective method of execution. A given plan includes selecting the best features for object detection and identification, selecting procedures for the data correction, determining the integration and classification procedures, and finally generating sub-plans to execute for improving the accuracy of the results. Section 4 describes the hierarchy of experts for this task.

3.2 Modelling Environmental Changes

Modelling of the scene and the objects is an important technique for integrating remote sensing, ground measurements of biophysical parameters, and GIS. As a focus of attention, we consider the monitoring of forests. The hierarchical forest model (Figure 1) is split into a forest growth model and a forest scene model. The forest growth model uses biophysical data, present forest cover, historical forest cover, and climate and hydrological projections in order to model the biomass and successional states of the forest. Physical measurements include climatological, hydrological and biochemical data while the ground measurements include tree species, tree characteristics (diameter, height, leaf area index, leaf moisture content, etc.), as well as soil measurements (soil type, moisture content, etc.).

The forest scene model incorporates the remote sensing data. The output of this model, run in the forward direction, is a simulated scene for a future time which can be compared with scenes acquired at that time. The forest scene model has three sub-models, a sensor model, a forest canopy reflectance model, and an atmospheric model. Components of the forest canopy reflectance model are the soil model, radiometric and geometric tree models and an understory model. The models may be run in an inverse mode in which remote sensing imagery is used to derive forest canopy modelling parameters. The tree model, for example, draws upon a spectral database for trees, tree components, soils, rocks, and the understory.

Combining the forest scene model with the forest growth model into a hierarchical forest model, we hope to be able to scale local measurements to larger scales. This approach or something very similar is what is required to relate remote sensing observations with predicted environmental states for various objects.

3.3 Assessing the Worth of Information

If an expert system is to evaluate and select the data for an information extraction computation, then the information content of the various data sources must be quantified. In general, it is very difficult to quantify the amount of useful information contained in an image. The usefulness of an image to a user depends on how many of the features in the image are directly related to his or her goals. Because the user's goals vary greatly, the information content or value of the same image or data source changes as a function of the user's goals. In order to quantify the information content, we would have to know the methods that can be used to achieve the goal (extract the desired information) and the amount of support which can be provided by a given data source to achieve the goal.

The value of a particular data source can be measured in terms of the amount of acquisition and computation costs to produce the needed support. The support of a goal for an image, for example, can be defined in terms of well known image descriptors such as linear features, regions, shape, local variance, moments, spatial and spectral frequency content, concavity, convexity etc. Odum (1988) has suggested the use of transformities as an energy-scaling factor for hierarchies of information. Given a search space of multiple routes to a desired goal, one selects that path from initial data to desired information goal requiring a minimum transformed energy. We are now exploring a variety of mathematical expressions and models for computing the optimal selection of data sources for a

given goal.

3.4 Data and Knowledge Representation

At CCRS, an hierarchical expert system, the Analyst Advisor, has been developed (Goodenough et al, 1988) for updating a forest GIS with objects derived from remote sensing imagery. The Analyst Advisor contains more than 30 expert systems built upon our shell, RESHELL. EMSys, described in greater detail in section 4, will also be an hierarchical system of experts. Each expert system in the Analyst Advisor contains rules with a specific scope of knowledge. The rules are obtained by extracting the knowledge of image analysts, foresters, and photo-interpreters. For environmental change detection it will be necessary to include knowledge from biologists, biochemists, meteorologists, ecologists, geologists, and a variety of resource disciplines. Each expert in the system determines which other experts it needs to consult with in order to meet its goal.

Knowledge and data representation will be in several forms. For efficient communication amongst experts, a frames database is used. This frames database stores names of physical objects and a list of associative information about each object. The associations include the object's relationship with other objects, and quantitative or qualitative information about the object. Knowledge is also represented in the form of production rules. Uncertainty values can be assigned for frames, objects, and rules. Data is stored in various formats, including grid, vector, polygonal, scalar, and symbolic. The expert systems transform the data into information through many processes. The highest level expert subdivides the user's goal into subgoals for each expert system. The responses of the expert systems and the associative relationships of the objects and frames are used to generate the information requested by the user. A research problem is the distribution of knowledge throughout the system. For example, does an expert system for segmentation with optical images require knowledge of forests or can the forestry knowledge be confined to higher level experts?

4.0 DESIGN OF AN EXPERT SYSTEM HIERARCHY FOR ENVIRONMENTAL MONITORING (EMSys)

4.1 Proposed Hierarchy of Experts

The structure of the expert system must take into account that several levels of processes and knowledge storage will occur. The lower level must include the more procedural and domain-specific experts. This will ensure each expert has a unique purpose. As one moves up the hierarchy, the expert systems make decisions based upon generalization of information and data from lower levels. Our shell, RESHELL, will not upload knowledge from a lower level expert system unless the expert is being consulted in the decision making process. This provides efficiency in memory management, modular development, and simplifies the maintenance of the expert system hierarchy. The RESHELL structure also allows each expert system to be also a network of expert systems, each one having a specific purpose and an ability to call upon other expert systems. It is this ability which allows us to build complex systems for multisource integration.

The structure in Figure 2 shows the proposed hierarchical expert system structure for environmental change detection, EMSys, using the philosophy described above. There are three major groupings: Goal Driven

Models, Environmental Assessment, and Data Analysis and Integration. The level entitled Data Analysis and Integration contains three parallel data interpretation domains: Field Data, Remote Sensing Data, and GIS data. Each data interpretation and analysis function is a network of expert systems. The data analysis level controls multiple experts to produce the features and analyses required to solve a particular goal. The higher level interprets the results of the data analysis within the data domain, and draws upon, when necessary, the inputs and outputs of the two other domains. The data integration level consists of a single expert system to integrate the three data domains, to resolve conflicting interpretations, and to pass symbolic knowledge and relevant measurements to higher levels. Conflict resolution may require new goals to be generated for the data interpretation experts in an iterative sequence.

The Environmental Assessment level applies physical meaning to the results of data integration. One expert system uses the integrated information to extract the environmental parameters of interest, while a second expert system reviews the past history of environmental change for the area. Both of these systems are used by an expert system which determines which changes have environmental significance. The Change Determination system also has access to the integrated data from lower levels. The Change Determination system will make extensive use of uncertainty values from previous interpretations. The top level (Goal Driven Models) draws on experts similar in function to Figure 1 to link environmental parameters in an environmental model, and to give predictions of future changes in a particular area. These expert systems are described in more detail in sections 4.2 and 4.3. Separating the expert systems into levels of scope does not preclude high level experts communicating to experts at a much lower level, if required.

4.2 Distribution of Knowledge over the Hierarchy

Our shell, RESHELL, allows for a network of communicating expert systems. A hierarchy of knowledge within the network is desirable in order to maintain a narrow focus for each expert system and to modularize the development process. The breadth of knowledge ranges from general semantic rules at the user interface level of the system to task-oriented, data manipulation experts at the lowest level. The partitioning of data types in the two lower sub-levels of the Data Analysis and Integration level allows the systems in these levels to maintain minimum scope. The analysis and interpretation experts for each data domain communicate through their common, higher level expert.

The Environmental Assessment level also has two expert systems which communicate through a higher level expert. The Environmental History Assessment expert detects trends in the data patterns over time, and feeds the results to the Change Determination system. The Extract Environmental Parameters system has knowledge of the physical meaning of the data. It will interpret the integrated data in terms of environmental parameters and supply relevant parameters to the Change Determination expert.

It is the Change Determination expert system which uses historical data and physical parameters to detect relevant environmental changes. All of this is done in the context of user goals. This expert may report directly to the user, or it may feed information and data to the Model Interpreter expert system. The Model Interpreter expert will be used in several different

ways. Since it can generate synthetic imagery corresponding to a particular environmental state, it will be useful in providing feedback to the interpretation experts and in developing probabilities for the various paths for interpretation. The Model Interpreter expert will also be used to generate predictions for various states which will feed into the Environmental Predictions expert. The Model Interpreter expert will also provide quality control on the lower level experts. The Model Interpreter will be consulted by the Environmental Prediction system to predict environmental change over time for various climatic scenarios. The Model Interpretation expert system is the most complex of the expert systems in the structure. Dealing with forest modelling, for example, it incorporates experts in forest growth models (which includes biomass models and succession models), forest scene models (including models for sensors, forest canopy reflectance and the atmosphere). These models require data, parameters and knowledge from other levels in the hierarchical structure, such as imagery and GIS files, climatic and hydrological parameters, and change probabilities derived from past assessments.

Any user interface to the environmental expert system must accommodate different types of queries and demands of the user on the system. User queries may range from "How much spruce do I have in my area" to "How much forest depletion can be expected in the next five years in my area due to acid rain?" The first question requires only a simple expert to review the GIS data and report the result. The second question requires a model interpretation of the area with knowledge about the level of acid rain in the area, the history of forest degradation in the area, how much of the previous degradation is due to acid rain, and the extrapolation of this into the future. For these reasons, the user interface expert system will require sufficient knowledge of the structure of the entire network to produce a complete agenda based on a user query. In order to respond to the range of queries, the user interface must communicate with all levels of the expert networks.

4.3 Application of the Expert System Hierarchy

This paper deals primarily with environmental change detection using forest monitoring from remote sensing platforms. Extensive and continuing collaboration with the British Columbia Ministry of Forests (BCMF) has provided experience in forest change detection using remote sensing (Goodenough, 1988). The resulting expert system network, the Analyst Advisor, has been written at CCRS with an emphasis on forest mapping for use by BCMF. The Analyst Advisor is embedded in the Data Analysis and Integration level of EMSys.

EMSys will initially be applied to boreal and alpine forests, since these are biomes widely distributed and sensitive to climatic change. The application of EMSys will require a dynamic, evolving development process. Much of the knowledge required for global monitoring does not currently exist. The modular approach proposed here should provide a framework in which to embed this knowledge as it is acquired. It is relatively straight-forward to provide tools which enable the participating experts to put knowledge (rules, frames, and objects) themselves into EMSys. More difficult is the truth maintenance across such a network of experts and the obvious requirement to have a learning capability. In the Analyst Advisor, we have already solved the problem of running multiple

experts on several computers communicating electronically.

5.0 CONCLUSIONS

The problem of environmental monitoring has been described as well as the importance of multiple source integration in predicting future environmental states. In an area where there are many experts, a complex search space for possible solutions, and uncertain data, information and knowledge, it has been proposed that artificial intelligence, initially through expert systems, is an essential technology to integrate multiple sources of images, maps, and ground measurements with the knowledge of diverse experts. Since there are many analysis paths, there is a requirement for a quantitative method of assessing the value of a data source. We propose that this can best be accomplished by knowing the costs to transform n-tuples of data sources into the information needed to answer a user's query. This analysis is followed by a high-level design (Figure 2) of an environmental monitoring system, EMSys, built upon an existing network of multiple experts for forest monitoring.

ACKNOWLEDGEMENTS

We wish to acknowledge valuable conversations with the following individuals: Frank Hegyi (BCMF); Phil Teillet, Karl Staenz, Robert O'Neil, Josef Cihlar, and Alain Menard of CCRS; and attendees at NASA's Boreal Forest Workshop in Durham, New Hampshire, December, 1988.

REFERENCES

- Parry, M. L., T. R. Carter, and N. T. Konijn (ed.) 1988, The Impact of Climatic Variations on Agriculture, Volume 1: Assessment in Cool Temperate and Cold Regions, Kluwer Academic Publishers (Boston).
- Goodenough, David G. 1988, "TM and SPOT Integration with GIS", Journal of Photogrammetric Engineering and Remote Sensing, 54, No. 2, pp. 167-176.
- Goodenough, David G., Bruce Baker, Gordon Plunkett, and Dena Schanzer 1989, "An Expert System For Using Digital Terrain Models", Proc. IGARSS'89, (in press).
- Goodenough, David G., M. Goldberg, G. W. Plunkett, and J. Zelek 1987, "An Expert System for Remote Sensing", IEEE Trans. on Geosciences and Remote Sensing, Vol. GE-25, No. 3, pp. 349-359.
- Odum, Howard T. 1988 "Self-Organization, Transformity, and Information", Science, 242, 1132-1139.

HIERARCHICAL FOREST SCENE AND GROWTH MODEL STRUCTURE

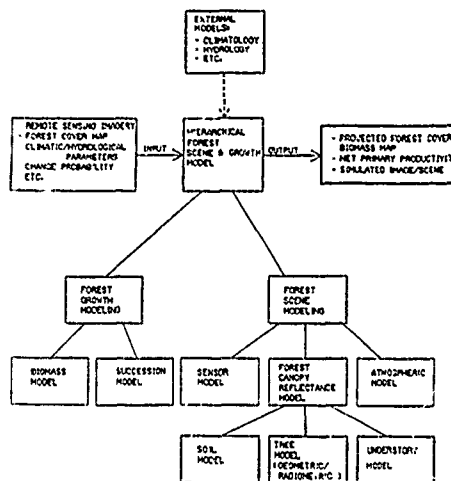


Figure 1. This figure shows the modeling network functioning in the forest scene and growth model. The forest growth model deals with the tree biology and ecology. The forest scene model deals with all factors affecting a remotely sensed image and includes the ability to simulate imagery from knowledge of ground cover.

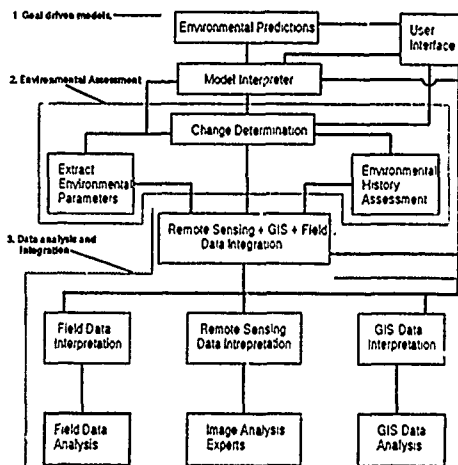


Figure 2. The expert system structure for EMSys shows three expert levels in the network. The bottom level contains the experts for data analysis, interpretation, and integration. The middle level extracts environmentally relevant information from the data, while the top level fits the environmental information; the models and makes predictions. The user interface links to all levels.

MULTISOURCE DATA ANALYSIS IN REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEMS BASED ON SHAFER'S THEORY OF EVIDENCE

Hakil Kim and Philip H. Swain

School of Electrical Engineering
and

Laboratory for Applications of Remote Sensing
Purdue University, West Lafayette, IN 47907 U.S.A.
Tel: (317) 494-1742, FAX: (317) 494-6440

Abstract

While it is empirically reasonable to assume that multispectral data have the multivariate Gaussian distribution, geographic or topographic data combined with multispectral data may not be represented by any parametric model. Furthermore, there is a difficulty in describing the various data types which have different units of measurements. These problems have been the motivation for the development of classification techniques in which various sources of data are assessed separately, and individual assessments are combined by some means.

In this paper, we present a method for multisource data classification based on the Shafer's mathematical theory of evidence. In this method, data sources are considered as entirely distinct bodies of evidence providing subjective probabilistic measures to propositions. In order to aggregate the information from multiple sources, the method adopts Dempster's rule for combining multiple bodies of evidence. The focuses of the paper are on 1) construction of support functions (or plausibility functions) given a body of statistical evidence, and 2) inferencing mechanisms of Dempster's rule in combining information under uncertainty.

Preliminary experiments have been undertaken to illustrate the use of the method in a supervised ground-cover type classification on multispectral data combined with digital elevation data. They demonstrate the ability of the method in capturing information provided by inexact and incomplete evidence when there are not enough training samples to estimate statistical parameters.

Key words : classification, combining function, evidential reasoning, multisource data, remote sensing.

1. Introduction

During the last decade, as remote sensing and other data acquisition technologies have advanced, there has been a trend toward exploiting remotely sensed multispectral data in conjunction with related data from other sources for the purpose of extracting higher level information from multi-attribute data bases. For instance, the topographic information obtained from digital terrain data has been successfully used together with remotely sensed data in land cover analysis (Franklin *et al.*, 1986; Jones *et al.*, 1988; Strahler *et al.*, 1978). More recently, a lot of researchers in the geographic information processing community have started reconsidering the possibility of utilizing remotely sensed data within geographic information systems (GIS) (Healey *et al.*, 1988; Quarmby *et al.*, 1988). With the advancement in designing sensor systems and the increasing availability of ancillary data, interest in extracting the great wealth of higher level information contained in geographic and remote sensing contexts has led to extensive demand for computer-based, automated (or semi-automated) methods for the analysis of multisource data. Their development will be hastened more and more by proliferation of various and sophisticated remote sensing platforms and sensors in the next decades.

Unlike the situation where we are dealing with purely spectral data from a single sensor, there are some significant factors which must be taken into account in devising means for multisensor and multisource data analysis, such as diverse data types, unknown interactions among sources, and unequal reliabilities of disparate sources. These problems have been the motivation for the development of the techniques by which inferences can be drawn systematically from complex data bases composed of disparate, unequally reliable sources, regardless of their data types and interactions with the other sources.

There have been a number of different approaches to the analysis of multisource data in remote sensing and GIS. First of all, the "stacked vector" approach is the most straightforward method, in which all data sources are considered simultaneously by organizing the respective measurements into a single vector, and the compound vectors are treated as data from a single source. However, its use is limited to the situation where the sources are similar and their interactions are easily modeled.

The "layered" approach proposed by Fleming *et al.* (Fleming, 1979) is more general in the sense that it can treat multiple sources of diverse data types separately. Hutchinson (Hutchinson, 1982) has developed a similar one, so called, "ambiguity reduction" method, whose basic strategy is to stratify the data based on one (or more) of the data sources, assess the results, and resort to the other sources to resolve the remaining ambiguities. A major disadvantage of these two approaches is that different groupings or orderings of the sources produce different results. In their schemes, the reliabilities and interactions of the sources cannot be incorporated systematically in the process.

Swain *et al.* (Swain, 1985) proposed an approach which can handle an arbitrary number of independent data sources. In their mathematical framework, the global membership function is derived from Bayes' formula by applying a statistical independence assumption. Due to the commutative property of the global membership function, different orderings of the sources in combination do not have an effect on final results. More recently, this method has been extended by Lee *et al.* (Lee, 1987) so that the relative quality of the sources can be embedded in the global membership function.

Although their combination procedures are different, the numerical representations of information in the above approaches are commonly based on the parametric Bayesian inference. It is very important to recognize that in dealing with multispectral data combined with other forms of geographic data, the methods employed must be able to cope with uncertainties which arise both from intrinsic randomness of data and from ambiguities in modeling and combining disparate sources.

The objective of our research is to develop a general, computer-based method of classification for multisource data in remote sensing and GIS. The method presented in this paper is based on the Shafer's theory of evidence, where the body of evidence provided by each data source is represented by non-additive probabilities rather than conventional additive probabilities so that uncertainty can be included as a measure. In order to aggregate the information from multiple sources and to propagate the uncertainty throughout the combination of information, this method employs Dempster's rule of combination.

The primary focus of this paper is on the construction of support functions (or plausibility functions) given a body of evidence. Secondly, Dempster's rule for combining evidence is examined in the sense of the desirable properties which agree with human intuition.

Viewing the problem of multisource data classification from the standpoint of evidential reasoning, we can present a way of modeling human reasoning under uncertainty in pattern recognition and data classification, where the information sources or decision rules are only partially reliable.

2. Constructing Support Functions In Shafer's Theory

One of the fundamental problems in applying Shafer's theory to real-world problems is how to construct either support functions or plausibility

functions given a body of evidence. In fact, this problem occurs in any other theory of subjective probability.

Shafer (Shafer, 1976) describes how to build the linear plausibility function when a given body of evidence is provided by a statistical experiment. In this section, after examining the characteristics of the linear plausibility function, we formulate a more general plausibility function by defining "partial consonance" (Walley, 1987).

Let Ω and X respectively denote a frame of discernment and a space of the observed data in a statistical experiment. Suppose that the observations are governed by a probability model $\{A_\omega, \omega \in \Omega\}$, where A_ω is a conditional probability density function on X given ω . The linear plausibility function based on this body of statistical evidence is derived from the following two assumptions:

- 1) the degree of plausibility of a singleton $\omega \in \Omega$ is proportional to A_ω
- 2) the plausibility function is consonant.

The first assumption corresponds to our intuitive feeling that an observation $x \in X$ favors those elements of Ω which assign the greater chance to x . Thus, x should determine a plausibility function P_x obeying

$$P_x(\{\omega\}) = C \cdot A_\omega(x) \quad \text{for all } \omega \in \Omega \quad (2.1)$$

where C is a constant which does not depend on ω .

A plausibility function $P: 2^\Omega \rightarrow [0, 1]$ is said to be consonant in Ω if it satisfies the following condition:

$$P(A \cup B) = \max\{P(A), P(B)\} \quad \text{for all } A, B \subset \Omega \quad (2.2)$$

The first assumption, together with the second assumption, determines a unique consonant plausibility function

$$P_x(A) = \frac{\max_{\omega \in A} P_x(\{\omega\})}{\max_{\omega \in \Omega} P_x(\{\omega\})} \quad \text{for all } A \subset \Omega \text{ and } A \neq \emptyset \quad (2.3)$$

When A is a singleton, say $\{\omega\}$, the consonant plausibility function gives the relative likelihood of ω to the most likely element in Ω . The corresponding consonant support function S_x is given as:

$$S_x(A) = 1 - P_x(\bar{A}) = 1 - \frac{\max_{\omega \in \bar{A}} P_x(\{\omega\})}{\max_{\omega \in \Omega} P_x(\{\omega\})} \quad \text{for all } A \subset \Omega \quad (2.4)$$

The linear plausibility method described above is simple to implement, and the principle behind it is similar to the Likelihood principle. However, its application is limited to the particular cases where the consonance assumption is satisfied.

A generalized scheme of the linear support and plausibility functions can be formulated by weakening the consonance assumption. The following definition of partial consonance originates from Walley (Walley, 1987) and is differently stated in this paper.

Definition. A support function $S: 2^\Omega \rightarrow [0, 1]$ is "partially consonant" iff there exists a partition $\{W_1, \dots, W_r\}$ of Ω and S is consonant in every W_k , $1 \leq k \leq r$, i.e.,

$$S(A \cap B) = \min\{S(A), S(B)\} \quad \text{for all } A, B \subset W_k \quad (1 \leq k \leq r).$$

Partial consonance allows a body of evidence to be heterogeneous, i.e., to support Ω in several directions. Consonance is the special case of partial consonance defined above when $r = 1$, i.e., the partition of Ω is Ω itself.

The partially consonant support function based on the body of statistical evidence is given as:

$$S_x(A) = C_p \sum_{k=1}^r \left[\max_{\omega \in W_k} P_x(\{\omega\}) \cdot \max_{\omega \in A \cap W_k} P_x(\{\omega\}) \right] \quad \text{for } A \subset W_k \quad (\text{some } 1 \leq k \leq r) \quad (2.5)$$

where C_p is a normalizing constant obtained as:

$$C_p = \left\{ \sum_{k=1}^r \max_{\omega \in W_k} P_x(\{\omega\}) \right\}^{-1} \quad (2.6)$$

The corresponding plausibility function is given as:

$$P_x(A) = C_p \sum_{k=1}^r \max_{\omega \in A \cap W_k} P_x(\{\omega\}) = \sum_{k=1}^r \max_{\omega \in A \cap W_k} P_x(\{\omega\}) \quad (2.7)$$

While partial consonance gives a flexibility to the linear plausibility function proposed by Shafer, it raises another problem of finding an optimal partitioning. The validity of its use is still under investigation.

3. Inferencing Mechanisms of Dempster's Rule

To base inferences and decisions on all available information in multi-source data analysis, it is necessary to combine the information from various sources. The role of rules for combining evidence is to integrate the conditional knowledge about states of nature based on each single body of evidence into total knowledge based on the combined evidence.

Several subjective Bayesian updating rules have been developed by applying one or two statistical independence assumptions to Bayes' rule. They have been successfully used in rule-based inferencing systems such as MYCIN (Shortliffe, 1976) and PROSPECTOR (Duda et al., 1979). However, there have been some controversies over the inconsistency between the independence assumptions and their updating rules (Glymour, 1985; Johnson, 1986; Pednault et al., 1981).

Dempster's rule is a generalized scheme of Bayesian inference to aggregate evidence provided by disparate sources. In this section, we examine Dempster's rule in the sense of the inferencing mechanisms.

Suppose that m_1 and m_2 are the basic probability assignments based on entirely distinct bodies of evidence, E_1 and E_2 . For all $A_1, A_2, A_k \in 2^\Omega$, Dempster's rule produces a new basic probability assignment m as:

$$m(A_k) = (1-K)^{-1} \sum_{A_1 \cap A_2 = A_k} m_1(A_1) m_2(A_2) \quad (A_k \neq \emptyset) \quad (3.1)$$

where

$$K = \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1) m_2(A_2)$$

The above equation is called the orthogonal sum of m_1 and m_2 . The denominator $(1-K)$ is a normalizing factor to compensate for the measure of belief committed to the empty set.

There are several points of interest regarding Dempster's rule. First, it requires that the basic probability assignments to be combined be based on entirely distinct bodies of evidence. Combining entirely distinct bodies of evidence may be considered as a fusion of the individual observations made by independent observers on the same experiment. The meaning of independence here is that one's observation does not have effect on any of the other's, which is quite different from the conventional independence definitions in probability theory. Secondly, it is both commutative and associative. Therefore, different orderings or groupings of evidence in combination do not affect the result. Finally, there is an additional piece of information provided by this rule. The constant K in equation (3.1) is the amount of the total probability that is committed to disjoint (or contradictory) subsets of Ω . It represents a measure of conflict between two bodies of evidence. When K is equal to one, it means that the two bodies of evidence are completely contradictory, and the orthogonal sum of their basic probability assignments does not exist.

Once the basic probability assignment m is obtained, the degrees of support and plausibility for all $A \subset \Omega$ are respectively computed as:

$$S(A) = \sum_{B \subset A} m(B) \quad (3.2)$$

$$P(A) = \sum_{B: A \subset B} m(B) \quad (3.3)$$

In the framework of Dempster-Shafer theory, the evidential information concerning a subset A of Ω can be represented by an interval-valued probability, $[S(A), P(A)]$. The width of the interval corresponds to the measure of uncertainty. The width, after the combination by Dempster's rule, is no larger than those of the intervals before the combination. This seems intuitively reasonable because the measure of uncertainty is expected to get smaller as we gather more information.

4. Experimental Results and Discussion

The linear plausibility function and Dempster's rule have been applied to the problem of ground-cover classification based on multispectral data in conjunction with digital elevation data. In this experiment, the decision is made according to the maximum plausibility in order to be comparable with the maximum posterior decision rule.

Table 1 describes the set of data sources for the experiment. The image in this data covers a forestry site around the Anderson River area of British Columbia, Canada. Source 1 consists of 4-band airborne multi-spectral scanner data in the visible region. Sources 2 and 3 are synthetic

References

1. Duda, R. J. Gaschnig, and P. Hart, "Model design in the PROSPECTOR: consultant system for mineral exploration", in D. Michie, (Eds.), *Expert Systems in the Micro-electronic Age*, Edinburg Univ. Press, Edinburg, pp157-167, 1979.
2. Fleming, M.D., and R.M. Hoffer, "Machine processing of Landsat MSS data and DMA topographic data for forest cover type mapping", LARS Technical Report 062879, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907, 1979.
3. Franklin, J., T.L. Logan, C.E. Woodcock, and A.H. Strahler, "Coniferous forest classification and inventory using Landsat and digital terrain data", IEEE Trans. Geos. and Remote Sensing, Vol. GE-24, No.1, pp139-149, 1986.
4. Glymour, C., "Independence assumptions and Bayesian updating", Artificial Intelligence, Vol. 25, pp95-99, 1985.
5. Johnson, R.W., "Independence and Bayesian updating methods", Artificial Intelligence, Vol. 29, pp217-227, 1986.
6. Jones, A.R., J.J. Settle, and B.K. Wyal, "Use of digital terrain data in the interpretation of SPOT-1 HRV multispectral imagery," Int. J. Remote Sensing, Vol. 9, No.4, pp669-682, 1988.
7. Healey, R., P. Dwyer, A. Mowle, and J. Holbrook, "Integrating remote sensing data into a geographic information system: A foundation for rural land use strategies - Nature conservancy council project", Proc. '88 IGARSS, Edinburgh, Scotland, pp111-112, Sep. 1988.
8. Hutchinson, C.F., "Techniques for combining Landsat and ancillary data for digital classification improvement", Photogrammetric Engineering and Remote Sensing, Vol. 48, No.1, pp123-130, 1982.
9. Lee, T., J.A. Richards, and P.H. Swain, "Probabilistic and evidential approaches to multisource data analysis", IEEE Trans. on Geos. and Remote Sensing, Vol. GE-25, pp283-293, 1987.
10. Pednault, E., S. Zucker, and L. Muresan, "On the independence assumption underlying subjective Bayesian updating", Artificial Intelligence, Vol. 16, pp213-222, 1981.
11. Quarmby, N.A., J.L. Cushnie, and J. Smith, "The use of remote sensing in conjunction with geographic information systems for local planning", Proc. '88 IGARSS, Edinburgh, Scotland, pp89-92, Sep. 1988.
12. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
13. Shortliffe, E.H., *Computer-Based Medical Consultations : MYCIN*, American Elsevier, 1976.
14. Strahler, A.H., and N.A. Bryant, "Improving forest cover classification accuracy from Landsat by incorporating topographic information", Proc. 12th Int. Symp. Remote Sensing of Environment, Environmental Research Inst. of Michigan, pp927-942, April, 1978.
15. Swain, P.H., J.A. Richards, and T. Lee, "Multisource data analysis in remote sensing and geographic information processing", Proc. 11th Int. Symposium on Machine Processing of Remotely Sensed Data, pp211-217, Purdue Univ., W. Lafayette, IN 47907, 1985.
16. Walley, P., "Belief function representations of statistical evidence", The Annals of Statistics, Vol. 15, No.4, pp1439-1465, 1987.

EXPERT SYSTEM DESIGN FOR RADAR RECONNAISSANCE OF SEA ICE

Glenn M. Shirtcliffe
 Radar Systems and Remote Sensing Laboratory
 University of Kansas, Lawrence, KS 66045
 Tel: 913/864-7735
 FAX: 913/864-7789
 Telex: 706352

Reliable information on the extent, thickness, age, and movement of sea ice is required for safe operation in Arctic waters. Synthetic aperture radar (SAR) has become a major tool in the operational reconnaissance of sea ice, but the logistics and practices of aerial SAR reconnaissance in the Arctic mitigate against widescale tactical support for the operational community.

Several SAR satellites are scheduled for launch in the next decade. These systems will provide the tactical support the operational community requires, but do so at very high data rates. The future problem will be one of data management. The volumes of data produced by these systems will strain the operational community's processing abilities, especially in the close-tactical time-frame.

The complex spatial and contextual information contained in SAR sea ice imagery requires an heuristic approach to analysis, more so than most other forms of image processing. A pixel-by-pixel approach to sea ice classification—the approach used in traditional digital image interpretation—ignores the syntactic (structural) and semantic (pragmatic) information that is present or inherent in the SAR image. Syntactic reasoning would, for example, suggest that if several points in a region share similar backscatter properties and are part of a linear structure, that they are part of a pressure ridge. Semantic reasoning would suggest that if one is operating in a tactical time-frame, one need only process the data of regional interest, and seek to identify in order: open water, leads, thin/new ice, thin first-year ice, and seek to avoid regions of pressure, ridging, glacial ice, or old ice. The pixel based approach of traditional image processing is most inefficient at using these higher forms of knowledge representation and reasoning.

Using the procedures and paradigms expressed in the field of artificial intelligence and symbolic image processing, research was conducted to investigate the application of such technologies to the interpretation and understanding of SAR images of sea ice.

A multi-resolution pyramidal data structure was selected to provide a hierarchical representation of image data for the system. The knowledge-source base and system control structure were implemented using object oriented programming techniques in the framework of a blackboard systems approach. The blackboard system structure itself is a hierarchical data structure that represents the problem domain as a hierarchy of analysis levels. The hierarchical structure of analysis and image data permits different analysis tasks to be performed at different levels of the blackboard or image pyramid in either a goal- or data-directed manner. Knowledge sources may be fired by goal postings or by data changes at pyramid levels above or below the current level of analysis.

The pyramid-based blackboard systems approach to image interpretation holds promise for a global image understanding system combining inputs from a multitude of sensors.

CLASSIFICATION OF MULTITEMPORAL REMOTELY SENSED
IMAGES BASED ON A FUZZY LOGIC TECHNIQUE.

P. Blonda (+), R. Loguercio Polosa (x), S. Losito (++),
A. Mori (x), G. Pasquariello (+), F. Posa (-), Ragno (o).

(+) I.E.S.I. - C.N.R., Bari (Italy).
(++) Centro di Geodesia Spaziale, Matera (Italy).
(x) Tecnopolis-C.S.A.T.A.-Novus Ortus, Valenzano (Italy).
(-) Dipartimento di Fisica, Bari (Italy).
(o) Istituto Biologico-Selviculturale, Viterbo (Italy).

ABSTRACT

A rule-based classification system for satellite multitemporal remotely sensed images, integrated with ancillary georeferenced data, is presented. As an example of non image data, the Digital Elevation Model (DEM) has been considered. The system knowledge-base consists of a set of rules describing each land-cover class for each date and with respect to the morphology. For the construction of these rules the system adopts, as user interface, linguistic descriptors such as: low, high, very low and any logical combination of them. These linguistic terms are treated as membership functions for "fuzzy sets" defined on spectral and altimetric values. The use of "fuzzy logic" is justified both by: i) the intrinsic uncertainty of the land-cover classes definition, ii) the difficulty of stating a phenological evolution in a large area with different micro-climatic and orographic conditions. The results of the rule-based approach are compared with a maximum likelihood supervised classification approach.

Keywords: Multitemporal Classification, Rule based knowledge, Uncertainty, Fuzzy possibility

I. INTRODUCTION

The objective of the present work is to define the initial requirements for a rule based classification system using TM or Spot imagery, integrated in a Geographic Information System, with the following characteristics: i) ease of man-machine interface, so that an expert agronomist could enter the system directly to define the knowledge base (K.B.) for each class; ii) applicability to the particular conditions of agriculture and land cover in Southern Italy, characterized (Caroppo, 1986) by an extreme fragmentation of the agricultural fields and a high variability of agricultural and land cover classes. The first one of these conditions makes the selection of spatial and contextual relationships very difficult; the second one demands for the definition of right decisional tools in the environment affected by high degree of indetermination

and uncertainty.

Furthermore, two aspects should be considered when a rule based tool is being used in conjunction with the spectral definition of classes: i) it is not always possible to define a value or set of values that are representative of a given class univocally; ii) the expert interpreter will use soft and imprecise expressions, typical of the natural language, to define the characteristic of a given class.

The parametric approach, traditionally used in the R.S. image classification, can provide an estimate of the indetermination level by means of random statistical distributions: the variability within a class is considered as noise superimposed to the real value (Kandel, 1978).

The present approach considers the "impossibility", in principle, of giving an univocal and deterministic definition for each real class. Hence the various classes are characterized only through the use of fuzzy possibility functions [(Zadeh, 1973), (Zadeh, 1974)]. The advantage is the availability of a straightforward user-system interface. In fact, the photo-interpreter makes large use of qualitative linguistic terms, such as high, low, the lowest, applied to the spectral response of a class, to describe the rules that associate the ground cover to the spectral characteristics of the image. These linguistic expressions can be considered as membership functions of fuzzy sets (Zadeh, 1973).

The adopted approach for the treatment of knowledge indetermination by means of "fuzzy" possibilities is presented in section II. The developed methodology has been applied to a Thematic Mapper image of Southern Italy. The test area characteristics and the selected classes are reported in section III and the applied knowledge base is described in IV. The obtained results of the classification, their discussion and the comparison with those obtained with a Maximum Likelihood approach are reported in section V.

II. THE FUZZY LOGIC AP: ACH

According to Zadeh definitions, given a set of values $U = \{x\}$, a fuzzy subset A of U can be defined as the set of pairs

$$A = \{ (x, mA(x)) \}$$

where $mA(x)$ is the membership function of A to U : it can assume real values within the interval $[0,1]$, expressing the measure for the "possibility" of x to belong to the subset A . In the case that $mA(x)$ assumes values equal to 0 or 1 only, the fuzzy subset definition coincides with the normal set definition.

Using the fuzzy set formalism, an interface language has been developed to describe the class spectral characteristics. This language is based on the definition of the following two membership functions:

$$LOW = \{ (x, ml(x)) \}$$

$$HIGH = \{ (x, mh(x)) \}$$

which represent the fuzzy subsets of "low and high spectral responses", respectively. Here, x is the pixel gray value in a given band, while the possibility functions $ml(x)$ and $mh(x)$ are defined as follows, according to the range X_{min} and X_{max} of the band under study:

$$ml(x) = \begin{cases} 1 - S(x, X_1, X_3) & , \text{ if } X_1 < x < X_2 \\ S(x, X_3, X_1) & , \text{ if } X_2 < x < X_3 \\ 0 & , \text{ if } X_3 < x < X_5 \end{cases}$$

$$mh(x) = \begin{cases} 0 & , \text{ if } X_1 < x < X_3 \\ S(x, X_3, X_1) & , \text{ if } X_3 < x < X_4 \\ 1 - S(x, X_1, X_3) & , \text{ if } X_4 < x < X_5 \end{cases}$$

where:

$$\begin{aligned} X_1 &= X_{min} & X_5 &= X_{max} \\ X_3 &= (X_5 - X_1)/2 + X_1 & X_2 &= (X_3 - X_1)/2 + X_1 \\ X_4 &= (X_5 - X_3)/2 + X_3 \end{aligned}$$

and

$$S(x, X_A, X_B) = 2 * ((x - X_A) / (X_B - X_A))^{**2}$$

This enables the use of the following operators contributing to the language assessment:

- 1) Enhancement: $VERY(A) \leftarrow (mA(X))^{**2}$
- 2) Dilution: $QUASI(A) \leftarrow (mA(X))^{**0.5}$
- 3) Negation: $NOT(A) \leftarrow (1 - mA(X))$
- 4) And: $A1 .AND. A2 \leftarrow \text{Min}(mA1(X), mA2(X))$
- 5) Or: $A1 .OR. A2 \leftarrow \text{Max}(mA1(X), mA2(X))$

From the operational viewpoint, an interactive environment is made available to the user, where he can set the definitions for the characteristics of the class C_i with respect to the generic band k of the GIS. This is accomplished through appropriate combinations of the following logical terms and operators:

(OUASI, VERY, NOT, AND, OR, LOW, HIGH).

Then the system translates any given definition in the corresponding function $m(C_i, k, x)$ automatically; each function represents the measure of the possibility to belong to the class C_i for a pixel with a value x in band k .

In order to handle multitemporal and

morphological data, the system allows the user to describe each class by means of only the "most useful" bands. A meta-level subsystem takes, for each class, the descriptions of the selected bands and submits these to the fuzzy classification subsystem, reducing, in this way, the system knowledge base to a well consistent one. If band k has been excluded by the meta-level subsystem for class C_j , the related function $m(C_j, k, x)$ is always set to 1 for each x . This indicates that every value in band k is compatible with the global class definition.

When the K.B. is defined, the system assigns each image point to a class for which the possibility function is maximized. The procedure is the following. Let:

$$X = \{x_1, \dots, x_k, \dots, x_n\}$$

be the set of the gray values for a given pixel in the n bands. The possibility for X to be included in the class C_i is defined as:

$$POSS(C_i, X) = \text{MIN} \{ m(C_i, k, x_k) \} , k=1, 2, \dots, N$$

[This choice insures that the most stringent requirements for the class selection are met; in fact the MIN corresponds to the AND operator]. Then, X is assigned to the class C_i if:

$$POSS(C_i, X) = \text{MAX} \{ POSS(C_i, X) \} \text{ when varying } C_i.$$

III. DESCRIPTION OF THE STUDY AREA

A test area of about 50,000 hectares has been selected, corresponding to a rectangular region of Southern Italy, with center coordinates $15^\circ E$, $41^\circ 20' N$ and heights ranging from 50 to 600 metres.

As large part of the Mediterranean agricultural context, the test area can be considered non particularly suitable from a remote sensing viewpoint, mainly because of the existing large variety in terms of land use classes and agricultural practices. In fact, the average size for homogeneous fields is less than 0.5 hectares and "mixed" agricultural practices, such as vineyards and olive groves or olive groves and croplands, are often used in the same parcel.

The study area is fully covered by the Landsat-D 31.188 imagery. Due to the complexity of cover classes, a multitemporal data set was employed, using three different dates in the year. The images were selected on the basis of the best data quality for each season. The spring image was acquired on April 86; the summer image on July and the Autumn one on October (the July image is shown in fig 1).

An integrated data set was obtained rectifying the images with respect to UTM projection for each date. A Digital Elevation Model was obtained digitizing contour lines from 1:50,000 topographic map, with 25 m. contour interval. The height error, computed as rms error on a set of fiducial points, was about 15 m.

The classes selection has been based upon information derived by: i) the latest Agricultural census (1982); ii) the study of topographic maps; iii) the local inspection of the area; iv) the visual interpretation of

related TM imagery both in single bands and RGB band superposition. The following classes have been selected:

- | | |
|----------------------------------|-----------------|
| 1) Bare soil | 5) Olive groves |
| 2) Urban areas | 6) Vineyards |
| 3) Pasture | 7) Cropland. |
| 4) Coniferous Reafforestation | |

IV. RULES DEFINITION

Rules have been assessed for the above classes in the three dates separately.

For the April date, the following relations exist: Bare soils are characterized by high and very high values in all the spectral bands; Urban areas show medium-high values; all the vegetated classes show low values in all the bands, but Pasture and Cropland have high values in band 4, because of high biomass content; the Olive groves characteristic is medium-low; for Vineyards, due to lack of foliage, the values are medium-high.

For the July date, Bare soils and Urban areas have high or very high values; Pasture and Cropland show medium-high values, but the latter are a little higher; low values can be found for those croplands already harvested; Vineyards, for its high biomass content, have low values in all bands, but high ones in band 4; Olive groves show medium-low values; Coniferous reafforestation has low or even very low values in all the spectral bands.

For the October image, the following relations can be stated: Urban areas show medium-high values in visible bands and medium in the infrared ones, while Bare soils are high or very high; the vegetated classes show medium values, with the following characterization: Coniferous reafforestations have lower values than the other classes, Vineyards still show high values in band 4; in this band the Cropland values are lower than Pasture ones.

It has been verified that the application of the full set of rules in the multitemporal approach, in some cases, leads to redundancy of characterization. This is particularly true for those classes showing a well-defined characterization in a few bands or a single date; in this case the meta-level subsystem allows to the expert to choose the reduced set of rules that is more suitable for each class.

From the morphological point of view, to describe the region of interest, we can introduce the following rules: Olive groves, Vineyards and Urban areas have a greater possibility for low altimetric values; Pastures and Coniferous are more favourite at not low altitudes. Within this region, the Cropland possibility is essentially unaffected by the morphology.

V. RESULTS AND DISCUSSION

In evaluating the described approach, special emphasis has been devoted to make a comparison between the results obtained using the fuzzy approach and those of a classical probabilistic approach. In order to achieve this goal, a supervised classification of the image data set has been performed.

A set of 25 homogeneous fields has been selected on the Landsat images and located on aerial photos at about 1:10,000 scale. The photointerpretation of selected areas was supported by ground investigation in order to establish an exact correspondence between training areas and ground classes. Using the training areas, a set of statistical features associated to each ground class was extracted. The whole set of statistical features was used in a straightforward manner, performing a Maximum Likelihood (ML) classification of the three temporal images as a single 18 spectral bands image.

The resulting map showed a high degree of overall correspondence with that one obtained by the fuzzy classifier: about the 65% of points in the maps had the same labels; this correspondence was more accentuated for pasture, olive groves and cropland classes.

In order to obtain a more quantitatively comparison between the two approaches, a set of 50 test areas was selected, for a total coverage of about 850 hectares, corresponding to the 1.5% of the global area. The distribution of the test areas versus the thematic classes agrees with the percentage of each class in the fuzzy classified map. Special care in the test areas selection has been devoted to find out the areas characterized by a larger difference between ML and fuzzy attributions.

A ground truth reference was obtained by photointerpretation and ground inspection of the test fields. Using this ground truth, a confusion matrix has been obtained for the fuzzy classification without using DEM data, as reported in table I.

As the table shows, the overall correct classification for the fuzzy system is about 91% versus the 86% value for the ML one. This situation is explained by a better attribution, in the fuzzy case, of urban, bare soil, coniferous reafforestation, and vineyards classes, whereas pasture, olive groves and cropland are classified with the same precision by the two systems.

Table II shows the final result obtained when applying the fuzzy rules to the data set integrated with DEM data. The corresponding classification map is shown in fig. 2. A slightly better value of overall classification accuracy (near the 94% of pixels are correctly classified) is obtained, because the residual confusion between Olive grove and Pasture classes is resolved using altitude data.

VI. CONCLUSIONS

In our opinion, when a scene shows a large percentage of non homogeneity, due for example to the fragmentation of the agricultural fields or to the variability in land cover classes, a non statistical approach, as a fuzzy based one, supplies with a better description of the real world and a classification system based on that description will give better results. In these situations, the improvement using a fuzzy description versus a statistical one will be directly proportional to the percentage of inhomogeneity in the scene.

References

1. CAROPPO, C., DI GENNARO, V., PASQUARIELLO, G., "UMUS: a demonstration project for usage of LANDSAT MSS and ancillary data in land cover mapping of large areas in Southern Italy". Proceeding of IGARSS'86 Symposium, Zurich, Sept. 8-11, 1986. ref. ESA SP-254.
2. KANDEL, A., BYATT, W.J., "Fuzzy sets, fuzzy algebra and fuzzy statistics". Proc. of the IEEE, vol. 66, pp.1619-1639, 1978.
3. ZADEH, L.A., "Outline of a new approach to the analysis of complex systems and decision processes". IEEE Trans. on Systems, Man, and Cybernetics, vol. SMC-3, pp.28-44, 1973.
4. ZADEH, L.A., FU, K., TANAKA, K., SHIMURA, M., "Fuzzy sets and their applications to cognitive and decision processes". Proc. of the U.S.-Japan Seminar on Fuzzy Sets and Their Applications, Univ. of California, Berkley, California, July 1-4, 1974.

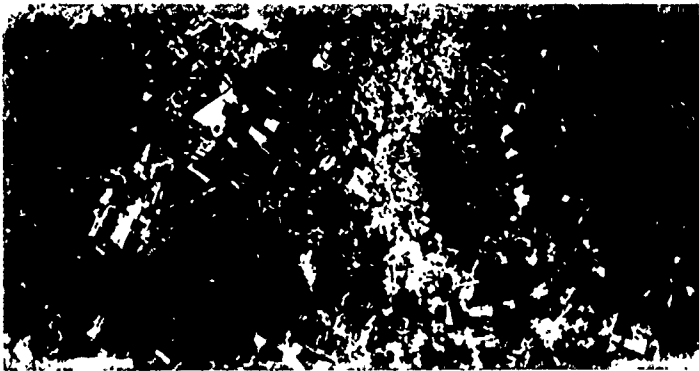


Fig. 1 - Landsat TM July image
(Bands 5,4,3 for RGB).

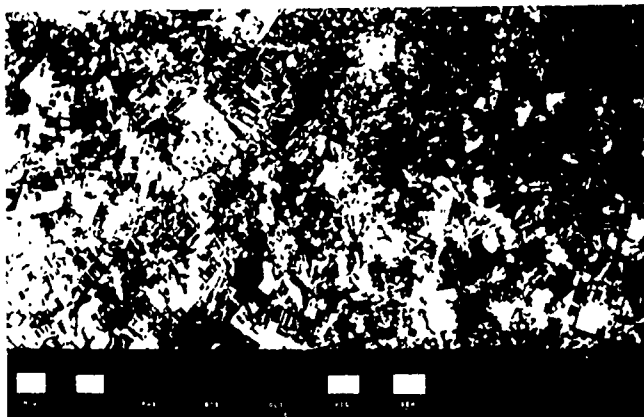


Fig. 2 - Fuzzy classification map.

GROUND TRUTH

| | BAR | URB | PAS | CON | OLI | VIN | CRO | TOT | PER |
|-----|-----|-----|------|-----|------|-----|------|------|-----|
| UNK | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| BAR | 211 | 1 | 4 | 0 | 49 | 3 | 1 | 269 | 78 |
| URB | 53 | 310 | 2 | 0 | 93 | 0 | 0 | 458 | 68 |
| PAS | 0 | 0 | 1311 | 0 | 3 | 0 | 165 | 1479 | 89 |
| CON | 0 | 0 | 250 | 166 | 9 | 0 | 0 | 425 | 39 |
| OLI | 1 | 0 | 90 | 2 | 2066 | 26 | 51 | 2236 | 92 |
| VIN | 0 | 0 | 0 | 1 | 121 | 644 | 99 | 865 | 74 |
| CRO | 0 | 0 | 8 | 0 | 22 | 0 | 1900 | 1930 | 98 |
| TOT | 265 | 311 | 1665 | 169 | 2363 | 673 | 2216 | 7662 | |
| PER | 80 | 100 | 79 | 98 | 87 | 96 | 86 | | 86 |

(A)

| | BAR | URB | PAS | CON | OLI | VIN | CRO | TOT | PER |
|-----|-----|-----|------|-----|------|-----|------|------|-----|
| UNK | 18 | 29 | 3 | 0 | 37 | 9 | 0 | 96 | |
| BAR | 224 | 14 | 0 | 0 | 0 | 0 | 0 | 238 | 94 |
| URB | 17 | 260 | 5 | 0 | 21 | 1 | 1 | 305 | 85 |
| PAS | 0 | 0 | 1480 | 1 | 146 | 27 | 28 | 1682 | 88 |
| CON | 0 | 0 | 3 | 162 | 9 | 0 | 0 | 174 | 93 |
| OLI | 6 | 8 | 158 | 0 | 2108 | 36 | 6 | 2322 | 91 |
| VIN | 0 | 0 | 0 | 0 | 8 | 598 | 3 | 609 | 98 |
| CRO | 0 | 0 | 16 | 6 | 34 | 2 | 2178 | 2236 | 97 |
| TOT | 265 | 311 | 1665 | 169 | 2363 | 673 | 2216 | 7662 | |
| PER | 85 | 84 | 89 | 96 | 89 | 89 | 98 | | 91 |

(B)

Table I
Classification results without DEM
(A: Maximum Likelihood; B: Fuzzy)

| | BAR | URB | PAS | CON | OLI | VIN | CRO | TOT | PER |
|-----|-----|-----|------|-----|------|-----|------|------|-----|
| UNK | 18 | 29 | 3 | 0 | 37 | 9 | 0 | 96 | |
| BAR | 224 | 14 | 3 | 0 | 0 | 0 | 0 | 241 | 93 |
| URB | 17 | 260 | 1 | 0 | 21 | 2 | 0 | 301 | 86 |
| PAS | 0 | 0 | 1586 | 1 | 67 | 13 | 24 | 1691 | 94 |
| CON | 0 | 0 | 4 | 162 | 9 | 0 | 0 | 175 | 93 |
| OLI | 6 | 8 | 51 | 0 | 2158 | 47 | 6 | 2276 | 95 |
| VIN | 0 | 0 | 0 | 0 | 8 | 599 | 3 | 610 | 98 |
| CRO | 0 | 0 | 17 | 6 | 63 | 3 | 2183 | 2272 | 96 |
| TOT | 265 | 311 | 1665 | 169 | 2363 | 673 | 2216 | 7662 | |
| PER | 85 | 84 | 95 | 96 | 91 | 89 | 99 | | 94 |

Table II
Fuzzy classifier results with DEM

INTEGRATION OF REMOTE SENSING AND GEOPHYSICAL/GEOLOGICAL DATA USING DEMPSTER-SHAFER APPROACH

Wooil M. Moon

Geophysics, The University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2

ABSTRACT

There are several methods available for integrating remote sensing data sets and also for integrating them with other information such as geophysical and geological data. Several published reports discuss successful application of different types of GIS (Geographical Information System). There have also been theoretical developments including Bayesian approach in updating old data sets with newly acquired information. However, there are still weaknesses and problems. Many geological and geophysical data sets often have only partial coverage and in almost all cases have very different spatial resolution. These cause serious difficulties in certain cases. In this research partial belief approach is examined as a mean to integrate remote sensing data with available geological and geophysical data and to successively update the existing information with newly observed data over target areas. In theory, Dempster-Shafer method appears to be the most suitable method but in practise there are several difficulties which have to be overcome. One of the major difficulties lies on dependency of the partial belief function on exploration targets, which can only be defined, at present, in case by case approach.

Keywords: Data integration, Dempster-Shafer method, Evidential belief function, Geological remote sensing.

1. INTRODUCTION

Integration of remote sensing and geological data has a long history, as old as the first geologist who tried to map outcrop rocks and draw the field information on a topographic map. This classical approach has been successful for simple tasks and required very little theory or research. However, with rapid advances in computer and space techniques, a digital GIS has been developed and the effectiveness of the powerful computer based GIS is generating renewed interest in remote sensing and earth science communities. The huge volume of remote sensing and geophysical data from the airborne and spaceborne platforms has also encouraged development of efficient GIS type methods of data integration. The most popular digital data sets beside the satellite images include geology map (lithological and structural), airborne magnetic (total field and gradient) map, gravity map (in less extent) and other geophysical data sets (Fig. 1). The available data sets are usually very limited for almost every target area, both in geological and geophysical maps and remote sensing image data. In this

paper GIS type approach of integrating remote sensing and other geoscience data will first be briefly reviewed and theoretical aspects of statistical and evidential belief function approach will be discussed.

2. GIS AND DATA INTEGRATION

There have recently been a number of papers dealing with GIS as a tool for geological and remote sensing data integration (Agterberg, 1988; Bonham-Carter et al., 1988; Bonham-Carter and Agterberg, 1989). Many of the basic functions in GIS such as preparation techniques of remote sensing and geological data inventory, base map preparation, thematic compilation and map integration, can directly be utilized in the geological remote sensing and geological data integration (Dangermond et al., 1987). Some of the technical problems involved in GIS systems, such as sliver errors, resolution inconsistencies and inconsistent map classes are now mostly resolved by efficient resampling and advanced interpolation techniques and optimized systematic integration approaches. However, there are further difficulties of thematic boundary variability, transition zones and questionable and missing data dilemma. There have also been some difficulties in vector and raster representation of data sets and transformation techniques.

These technical difficulties can be ironed out with evolution of the GIS system. The more serious problem appears to be precise representation of the information on each data set prior to data integration. In GIS, a given plane of data is generally reclassified or generalized before merging. Reclassification or generalization of attributes itself is not a technically difficult process (Dangermond et al., 1987). When it comes to reassigning of observed information on each data set, there exists considerable difficulty and misunderstanding on how each class should represent what range of signal power or mapped information. In most GIS systems, available today, points, lines and/or polygons represent either a certain information or feature or absence of them. In this approach any map information is represented as binary map of [0,1]. This approach poses serious problems for most geophysical survey data, where each contour interval represents a range of particular field values. Another difficulty arises from missing data. If a geological map, for example, shows only 10% outcrop distribution and the rest is covered by glacial deposits, integration of the basement geology with other data must have a formalism to represent the interpreter's ignorance or missing information.

In a recent development, called "weights of evidence" mapping, each input map is converted to binary form where the two map classes are determined by the target objectives (Bonham-Carter et al., 1988; Bonham-Carter and Agterberg, 1989). In the binary map analysis, for example, score for gold occurrence is set to 1 and non-occurrence to 0 and a GIS system such as SPANS* can be used to integrate geological, geophysical and mineral occurrence data and to evaluate a certain target potential (Bonham-Carter et al., 1988). (* - Trade Mark of TYDAC Technologies Inc). The weights of evidence modelling and binary map approach appeals due to its simplicity. However this method breaks down when the "weights of evidence" is required for data sets with continuously varying data values or for data sets which do not exist (unsurveyed or unexplored) (Zadeh, 1974; Bonham-Carter et al., 1988).

3. EVIDENTIAL REASONING AND DEMPSTER-SHAFFER METHOD

In the following discussions, digital map inventory preparation (geological and geophysical), creation of base map, resampling, interpolation and geocoding are assumed to have been carried out and emphasis will be focused on theoretical aspects only. It will also be assumed that each plane of information has originated from multiple but disparate sources. The information level or evidences can then either have varying degree of certainty to several environmental possibilities, a incorrect or are incomplete. If one can assign a degree of belief to each evidence, as in the evidential theory of belief, evidences with varying degree of certainty can be represented by partial belief functions (Shafer, 1976).

Suppose there is environmental possibilities e_1, e_2, \dots, e_n , such that

$$E = \{e_1, e_2, \dots, e_n\}$$

then each proposition is completely defined by the subset of E containing exactly those environmental possibilities where the proposition is true. If, in a given data set, geological formations e_i and e_{i+1} cannot be distinguished in terms of at least one proposition of interest, they should be replaced by a single environmental element. The proposition "an ultramafic stock is located at (x,y) " then corresponds to the subset of environmental possibilities, that some kind of ultramafic rock is located at (x,y) . Now a geophysicist can represent one's partial belief through a Bayesian distribution over E . This is done by distributing a unit of belief among the elements of E attributing commensurately greater amounts to the more likely elements. If one designates this distribution by the mapping 'dist'

$$\text{dist} : E \rightarrow [0,1]$$

$$\sum_{e \in E} \text{dist}(e) = 1.0.$$

This induces a probability on every proposition X defined over each layer of digital information (Fig. 1) such that

$$\text{for all } X \subseteq E, \text{ Prob}(X) = \sum_{e \in X} \text{dist}(e)$$

and it follows that

$$\text{Prob}(X) = 1.0 - \text{Prob}(\neg X).$$

The problem with this approach is that the geophysicist has to determine a precise probability for every proposition for each map layer no matter how impoverished the evidence. This would not be such a problem

if there were a rich source of statistical data for each map layer from which these probabilities could be estimated. Unfortunately, in most cases, each digital information layer of geological and geophysical data base is often incomplete and the Bayesian statistics tends to prefer disjunction of the mutually exclusive propositions which often results in unrealistic probability of certain propositions.

In the evidential belief function approach a proposition X is represented by an interval $\text{Spt}(X)$ and $\text{Pls}(X)$, where the latter represents degree to which the evidence supports the proposition X and the former represents degree to which the evidence remains plausible. This evidential interval is a sub-interval of the closed real interval $[0,1]$. The difference, $\text{Pls}(X) - \text{Spt}(X)$, represents residual ignorance of the given subset of environmental elements. If one has accurate and exact information, the evidential interval collapses to a point in $[0,1]$. In the binary map approach of integrating geological data (Bonham-Carter et al., 1988; Bonham-Carter and Agterberg, 1989) and many other data integration methods using GIS implicitly assume such hypothetical limits.

A unit of belief over a set of propositions can be distributed as a mass distribution where the focal propositions need not be mutually exclusive, such that

$$\text{mass} : 2^E \rightarrow [0,1]$$

$$\sum_{F \subseteq E} \text{mass}(F) = 1.0$$

$$\text{mass}(\emptyset) = 0.0.$$

Here, it should be noted that the sum of the mass attributed to propositions that imply X ($\text{Spt}(X)$) plus the sum of the mass attributed to propositions that imply $\neg X$ ($\text{Spt}(\neg X)$) do not necessarily equal 1.0, since some mass might be attributed to propositions that imply neither.

If rock formations represented by F could be included for the proposition X , an evidential interval can be induced on the probability of X such that

$$\text{Spt}(X) = \sum_{F \subseteq X} \text{mass}(F)$$

$$\text{Pls}(X) = 1.0 - \text{Spt}(\neg X)$$

$$= 1.0 - \sum_{F \subseteq \neg X} \text{mass}(F)$$

and

$$\forall X \subseteq E$$

$$\text{Spt}(X) \leq \text{Prob}(X) \leq \text{Pls}(X).$$

Viewed intuitively, mass is attributed to the most precise propositions a body of evidence supports. Bayesian approach requires that a precise probability be assigned to each evidence (eg. type of rocks), no matter how noisy the data (eg. uncertainty in identifying the rock types) are and no matter how little statistical data (eg. insufficient number of outcrops) are available. In Dempster-Shafer approach, one computes $\text{Spt}(X)$ and $\text{Pls}(X)$ based on an understanding of the propositional dependencies that exist within the environment under study. If a mass is attributed to some proposition X and it is not known whether X implies another proposition Y or $\neg Y$, then the judgement can be suspended and the integration process branches to an available vertical option. Mass(X) neither increases $\text{Spt}(X)$ nor decreases $\text{Pls}(X)$, but contributes to the evidential interval. This ability to represent the ignorance gives Dempster-Shafer approach a clear advantage over Bayesian approach which breaks down.

In Dempster-Shafer approach, Dempster's rule does not require that one body of evidence support a single proposition with certainty. The rule can take arbitrary complex mass distributions $mass_1$ and $mass_2$ and as long as they are not completely contradictory with respect to each other, and can produce a third mass distribution $mass_3$. Therefore, given F_1 , F_2 and F_3 in a digital map layer,

$$mass_3(F)_3 = \frac{1}{1-k} \sum_{F_1 \cap F_2 = F_3} mass_1(F_1) mass_2(F_2)$$

where

$$k = \sum_{F_1 \cap F_2 = \emptyset} mass_1(F_1) mass_2(F_2) < 1.0.$$

Since Dempster's rule is both commutative and associative, the order and grouping of combinations are immaterial. This fact allows results to be obtained through hierarchical combinations of partial results with whatever degree of parallelism, less depending on the nature of map layers.

In Dempster-Shafer approach, some information, as measure of conflict k , can also be provided in regard to gross error during the data integration. The measure of conflict k provides degree to which the combined information or the new compilation map is contradictory towards the propositions. Given several bodies of evidences, one can expect that those containing gross errors will tend to be farther away from the other bodies of evidence that those with measurement errors. One can use clustering algorithm to sort out those bodies of evidence containing gross errors.

One of the problems with Dempster-Shafer approach is that one must maintain each body of information independent because of the inability of the method to correctly combine information with known dependencies. Most of remote sensing and geophysical data sets are fortunately evidentially independent and problem of over-weighting one information or evidence does not require special attention (Moon, 1989).

4. DISCUSSION AND CONCLUSION

In Dempster-Shafer approach of integrating information, one must first be able to reason about possibilities and also about the interrelation of information. The human perceptual/reasoning system has already a multi-sensor data integration capability imbedded in it. With this capability an explorationist (geologist or geophysicist) can actively cue sensors, seeking confirmation and/or refuting evidences related to the exploration target entities.

In geological/geophysical remote sensing, most data sets are incomplete and statistically unbalanced even though the rate of increase in newly available data volume is alarmingly large. One approach of solving this problem will be a planned interactive approach of exploration, which is, at present, practically impossible. Dempster-Shafer approach provides an optimal theoretical basis for integrating remote sensing, geological and geophysical data sets. The straightforward approach of GIS technique, the binary map approach (Bonham-Carter et al., 1988; Bonham-Carter and Agterberg, 1989) and Bayesian approach of updating old data sets are all valid and also provide a tool but the choice will be determined by accuracy of the expected outcome and exploration objectives. One of the difficulties of applying Dempster-Shafer approach in geological/geophysical remote sensing lies in the variation of the evidential belief func-

tion which critically depends on the final exploration target.

5. ACKNOWLEDGEMENTS

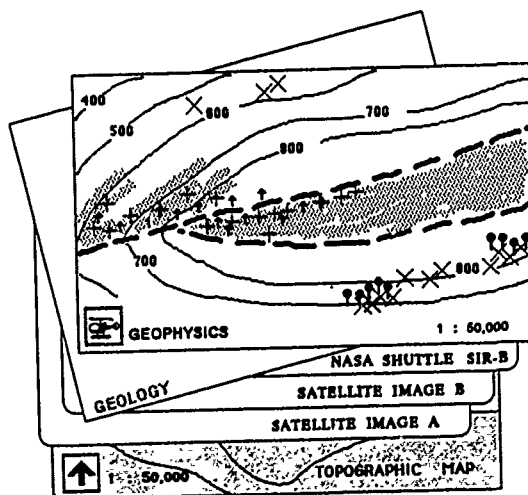
This research is supported by NSERC operating #A-7400 to W.M.Moon. Some of the ideas in this paper was conceived during the discussion with Ping An. Author would like to thank Dr. G.F. Bonham-Carter for the preprints and initial discussion.

6. REFERENCES

- Bonham-Carter, G.F. and Agterberg, F.P., 1989. Application of a micro-computer based geographic information system to mineral potential mapping, Micro-Computer in Geology (submitted).
- Bonham-Carter, G.F., Agterberg, F.P. and Wright, D.F., 1988. Integration of geological datasets for gold exploration in Nova Scotia, Photo. Eng. and Remote Sensing, 54, 1585-1592.
- Dangermond, J. and Freedman, C., 1987. Description of techniques for automation of regional natural resources inventory, inGIS for Resources Management; A Compendium, (Editor; Ripple, W.J.), 9-33.
- Moon, W.M., 1989. Application of evidential belief theory in geological geophysical data integration, Geophysics (submitted).
- Shafer, G., 1976. A Mathematical Theory of Evidence, Princeton University Press, New Jersey.
- Zadeh, L.A., 1979. On the validity of Dempster's rule of combination of evidences, UCB/ERL M79/24, 1-12.

7. FIGURE CAPTION

Figure 1. Schematic diagram of remote sensing and other geological and geophysical data to be integrated.



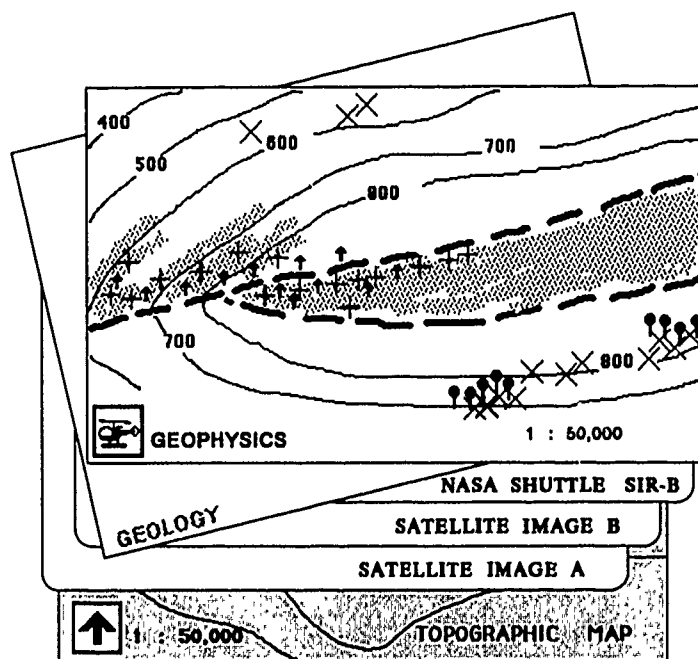


Fig. 1. Schematic diagram of remote sensing and other geological and geophysical data to be integrated.

AN EXPERT SYSTEM FOR USING DIGITAL TERRAIN MODELS

by

David G. Goodenough, Bruce Baker,¹
Gordon Plunkett, and Dena Schanzer¹

Canada Centre for Remote Sensing
Department of Energy, Mines and Resources
Ottawa, Ontario, Canada

In order to use high resolution, remote sensing imagery in mountainous terrain, one must make use of digital terrain models (DTMs). These models are used to provide elevation, aspect, and slope features for interpretation. They are used to geometrically correct the imagery for topographic relief. Such models are the basis for computation of shadowed areas and multiple reflectances. In simulating the image which would be produced for a given surficial cover, illumination angle, and viewing angle, one must use the DTM and the bi-directional reflectance functions of various covers. Given the extensive and complex uses to which DTMs can be put, it is of interest to know the impact of errors in the DTMs for these various uses. In order to simplify and codify this knowledge, expert systems can play an important role in advising users on the uses and restrictions for particular DTMs.

This paper reports on experiments conducted to investigate these errors. We have used elevation, slope, and aspect as features combined with spectral and textural features for Thematic Mapper imagery. DTMs have also been used to generate synthetic images corresponding to the solar illumination angles seen over one day (June 21) and over one year at the satellite imaging time. A multi-expert system has been developed to advise users on image analyst. An important component of the Data Preparation Expert is the DTM Expert. This expert system contains knowledge about errors found in DTMs. The reality of these errors is that they vary across the DTM and are functions of the methods used to make the DTM, the topographical relief, the surficial cover, and the positional location. This reality is modeled simply in the DTM Expert at this time. However, as we gain experience and knowledge of typical DTM errors, the DTM Expert will generate revised accuracy estimates for each point in the DTM. Depending upon the use of the DTM, it may be necessary to use the DTM at several different resolutions.

For example, for elevation errors at 1:20,000 scale, the computation of shadowed areas, for example, will have errors of no greater than 10m given an elevation RMS error of 9 meters (Lasserre²). A digitized 1:50,000 scale map (A1) with a contour interval of 30m can produce a DTM with 25m pixels, but to use such a DTM for slope and aspect calculations, it is necessary to consider interpolation errors as well and to degrade the

¹Employed by INTERA Technologies Ltd., Ottawa, Canada.

²Lasserre, M. 1989, private communication.

DTM resolution to perhaps 120m pixels (Peuker³) in order to have acceptable errors for slope, aspect, and other topographic descriptions. Geometric errors of greater than 0.5 pixels will substantially reduce the accuracy of a multi-dimensional classification.

The large angular size of the sun ($\frac{1}{2}^\circ$ arc) results in umbral and penumbral pixels. In other words, there are pixels completely in shadow and pixels partially shadowed. These pixels may receive reflectance contributions from opposite slopes which could lead to an incorrect interpretation of the objects.

The paper describes a Digital Terrain Model expert system which is integrated into a multi-expert system for advising on satellite image interpretation. This expert system is written in Prolog and controls processes written in FORTRAN on several computers. The expert system contains the knowledge derived from these experiments in the form of rules, frames, and procedures. The result is that the user is guided in the correct use of DTMs and thus avoids the common pitfalls of attempting to use DTMs at resolutions at which errors are large.

³Peuker, T. K., R. J. Fowler, J. J. Little, and D. M. Mark 1976, "Digital Representation of Three-Dimensional Surfaces by Triangulated Irregular Networks (TIN)," Tech. Report No. 10, Office of Naval Research, Arlington, VA, U.S.A.

A KNOWLEDGE BASED SYSTEM FOR THE INTERPRETATION OF SAR IMAGES OF SEA ICE

J.G. McAvoy and E. M. Krakowski

*INTERA Technologies, Ltd
Calgary, CANADA*

ABSTRACT

A knowledge based system for the interpretation of SAR images of sea ice has been developed to the prototype stage. The system contains approximately 100 rules and is capable of classifying SAR images based on a minimal set of feature information supplied by the user. The system runs on a small computer and can be used to assist non-experts make correct SAR ice interpretations. It also contains a SAR image database of representative ice types which can be displayed and used for training.

Keywords: SAR, Ice Classification, Expert Systems, Knowledge Based Systems

1. INTRODUCTION

A dominant factor in both Canada's offshore environment and other northern countries is the more or less continuous presence of sea ice. With the growing commercial development of these offshore frontier areas there has arisen the need for better, more reliable information on ice conditions. Those operators who require such ice information include: ice breakers and ice breaking tankers capable of year round operation, lightly reinforced vessels which have more restricted operations, and oil and gas drilling and production platforms.

Synthetic aperture radar (SAR) has proven to be an ideal sensor for mapping these ever changing ice conditions. Both airborne and spaceborne SARs typically achieve ground resolutions of the order of 5 to 30 metres which is suitable for mapping the important surface characteristics of sea ice under all weather and ambient light conditions.

It is possible to determine ice type, hence infer ice thickness using both SAR and a knowledge of the processes that act on an ice sheet

during its growth and development. One of the main problems with this procedure however, is the fact that an expert is usually required to make these interpretations.

This paper describes the development of a knowledge based system for the interpretation of SAR images of sea ice. The system has been developed to the prototype stage and demonstrates that an expert system approach can be applied to this and other complex image interpretation problems. Developing the system has also resulted in a much better understanding and structuring of the underlying knowledge required for the job of ice classification.

The system described is capable of classifying SAR images based on data input from users. It runs on a small computer and can be used in a stand alone mode, say on the bridge of a ship, to assist novices make ice interpretations. It also contains a digital SAR image database of representative ice types which can be displayed and therefore used as a training system by observing the way the system classifies.

2. KNOWLEDGE ACQUISITION

It was clear from the start of this program that the specific classification of ice types and the features that contribute to the definition of these types is a subject of some controversy and is not well documented. It is further complicated by the fact that experts often disagree about the importance of certain features required for classification particularly when classifications are primarily made on the basis of a single sensor like SAR.

INTERA's initial interviewing and questioning of several ice classification experts revealed that it was very difficult for them to describe what they do when they classify. Experts generally have no model upon which to base their expertise. A series of structured experiments were therefore undertaken to elicit this knowledge using

an image display system and a computer based knowledge engineering tool developed by Shaw and Gains [1] which is based on Kelly's Personal Construct Psychology [2].

The knowledge engineering tool called PCS is an interactive computer aid which incorporates techniques for eliciting the expert's vocabulary and structuring some aspects of the domain knowledge. The methodology employed involved six experts, making comparisons between a triad of randomly chosen SAR images, viewed in sequence, in order to describe similarities and differences. Subjects completed 20 trials and were asked to indicate: how a given image was different from the other two images observed, how the other two images were similar, and to generate a construct name for the features etc. that distinguished these images. Subjects then rated each image with respect to this construct using a five point scale to indicate the presence or absence of a particular feature. Subjects were also asked to rate the importance of each construct to the task of ice classification. Examples of constructs include the presence or absence of such features as: angular floes, fine texture, homogeneous tone, ridging, etc.

A repertory grid was constructed as a way of representing these constructs which contained a set of bipolar distinctions made about the images. The PCS program contained several features to analyse and determine inter-relations between constructs using distance-based cluster-analytic techniques, principal components methods, and Shaw's FOCUS. Constructs from all experts were pooled into an initial set of 31 constructs.

The second stage in the knowledge acquisition phase involved the establishment of a construct matrix where the subjects were asked whether a particular pooled construct was present or absent on a given set of images and to indicate the importance of the constructs to each ice type. The images included examples of the following ice types: New Ice, First-Year Ice, Second-Year Ice, and Multi-Year Ice. The results of this session showed that experts strongly agreed on the ice type when certain key constructs are either unequivocally present or absent.

These experiments generated a preliminary knowledge base of features and constructs which are used by experts for classification. It was clear however, that experts use a great deal more information when classifying, including knowledge of the regional ice climatology. The preliminary knowledge base was therefore augmented with additional knowledge extracted from two key experts using additional example images and an ongoing interviewing technique.

It should be noted that the image data sets used in these experiments

were limited to the ice conditions of the American and Canadian Beaufort Seas and the Canadian Arctic during freezing and melting conditions. Thus the domain of knowledge contained in the system is restricted to these areas.

3. KNOWLEDGE REPRESENTATION

A rule based system was chosen for the prototype since the classification procedure naturally seemed to fit into this form of representation. The rule sets were written for specific ice types and rely on the presence of image features and elements, as well as other information, such as time of year and geographic location. Only affirmative constructs are used for the "ice-type rules". These "affirmative" rules require that the appropriate constructs or features be present in the image. An example of an affirmative rule is:

If
there is evidence of floe
And floe.composition is "Conglomerate"
And floe.shape is "Round", "Round-Angular"
And tone.inner is "Grey", "Bright Grey", "Bright"
Then Old_Ice_3 is confirmed

In addition to affirmative rules a second type of rule termed "supportive" is also used. These supportive rules evolved out of the process of determining unique characteristics for each ice type, a task which proved to be rather difficult. There are however, several characteristics for each ice type which typically lend support to the interpretation, but which are not unique to that ice type. These are termed supportive constructs since they lend support to the conclusion reached via the affirmative rules. Once the affirmative rules are fired and a conclusion reached, the system checks the supportive rules to verify that this conclusion is correct. These supportive rules are also particularly valuable in training situations when expressed along with the affirmative rules since they show the user how the system arrived at its classification. An example of a supportive rule is:

If
there is evidence of floe
And there is evidence of floe.size
And there is evidence of ridge
Then Old_Ice_3 _support is confirmed

One more rule type has also been incorporated into the system. This rule type, termed "precluder" is used to exclude ice types whenever possible. The reasoning here is that there are times when an ice sheet has so few features that a classification is almost impossible. It is however, usually possible to exclude some ice types, which is in

itself, useful for the classification process. An example of an precluder rule is:

If
 there is evidence of rafting
Then not first year is confirmed

The above rules have not been designed to handle mixtures of different ice types since this was seen as a future development. It is interesting to note that the prototype runs in an exhaustive manner searching the knowledge base for matches until all rules have been checked. If the area of interest actually contains more than one ice type, the system will output results indicating that different ice types are present.

One addition to the system has been the division of the Canadian Arctic into four unique ice zones. These zones were determined from the identification of unique ice characteristics identified on numerous images. An example of one of these zones is the Northern Arctic Islands. Within the zone are three landfast multi-year ice plugs. By having the user identify the geographic area the system is more quickly able to determine the ice type.

4. PROTOTYPE SYSTEM

The prototype ice classification system has been developed using Neuron Data Inc's. expert system shell, Nexpert Object. Nexpert Object was chosen for the prototype since it is a relatively low cost tool written in "C" which can run on a variety of PC's, workstations, and mainframe computers. Thus, applications developed on one computer system can be easily ported to other systems. Nexpert Object is a hybrid tool, combining structured representations and rules. It is built around a reasoning kernel with forward and backward chaining, automatic goal generation, multiple inheritance, and direct calls to external routines. It also has an excellent developer's interface.

The system was developed on a 68030 workstation (MacIIx) containing 4 Mbytes of RAM, a 160 Mbyte internal hard disk and a high resolution (1152x882) colour monitor.

Nexpert Object also offered a unique feature: the ability to automatically create a visual representation of the links between concepts and rules (i.e. a rule network). This was an invaluable tool for the development of the prototype since it allowed both the system's developers and the experts to see a graphical representation of the rules and to trace the reasoning of the system during processing.

Both forward and backward chaining are employed in the prototype. The system processes rules on the basis of hypotheses suggested and from user input. The system proceeds with a depth first search from a root context to a context selected. A series of parameters, stored in meta-slots, have been customised to generate menus for user input whenever input is required during the inferencing process.

A custom graphics user interface was written for this application. This interface allows both text and imagery to be combined on the display. Digital images that are stored on the system are displayed in two windows. One window displays an overview image of wide areal extent. This overview image was found to be important since it contains the contextual information necessary for classification. A second window displays a zoomed-in portion of the overview image at full resolution. A coloured box is used to outline the area of interest and focus the user's attention on a given ice type. A third window located at the bottom of the screen is used for user interaction. The style of this interaction is based on the Macintosh interface standard where the user interacts with the system via a mouse and in some instances with an attached keyboard. A series of tri-state buttons are used to represent menu choices.

An important feature of the user interface is the use of templates. These templates are used to solicit information on criteria such as tone and texture where it is difficult to express in words a relative value for these attributes. For example, the user is presented with a grey scale and asked to match the overall tone of the image. Similarly a series of examples of various textured images are also presented and the user asked to match the image to one of the examples.

An extensive help facility is also incorporated into the system. The meaning of various terms used by the system is displayed automatically any time the user moves his cursor over a referenced term. In addition, actual image examples of various features and terms can also be displayed by hitting the help button.

It should be noted that the prototype development went through several iterations and evaluations before the current system emerged. This process of prototyping and evaluation proved to be an excellent way of soliciting further information from the experts and made it easier for them to communicate with the system programmers.

5. SYSTEM OPERATION

The prototype system has been designed to analyse any SAR image in the geographic regions outlined earlier. The system contains a number of SAR images, stored in digital format on disk that may be

brought to the screen for analysis. In addition the user may analyse other imagery which may be in the form of positive prints, film negatives, etc.

Users are asked to view an image and are then asked to input easily recognisable features, patterns, constructs, and other relevant information which is used by the system for classification. The area to be classified on the stored images is outlined and represents a single ice type. The system classifies the ice into one of the following types: New, Young, First-Year, Second-Year, and Multi-Year. First-Year ice is further divided into smooth, ridged, rough, and very rough.

In many cases the system is able to classify the ice with a minimum amount of supplied information, sometimes requiring only three inputs from the user. This has implications for those researchers attempting to perform machine classification since it may present a way of reducing the amount of processing required. In order to make the prototype more useful as a training tool, it was configured so that the user must go through a complete line of logical questioning before any output is presented. These questions are presented in an order thought to be a reasonable representation of the way that an experienced ice interpreter would proceed.

The output from the system consists of a display of the rules that are fired and the classification made. The rule types include the affirmative, supportive, and precluder rules outlined earlier. In order to make the system more useful for training, a journal file is also created on the system. This file is used to record the path of decision making that the system took to arrive at a classification.

A great deal more work is required to validate the accuracy of the classifications made by the system. At present it has been confirmed by at least one external expert that the system accuracy is about 75% for new, young, and first-year ice and about 85% accurate for old ice.

6. CONCLUSIONS

A knowledge based system for the interpretation of SAR images of sea ice has been developed that is capable of classifying SAR images based on a minimal set of feature information supplied by the user. The system can be used to assist non-experts make correct SAR ice interpretations and can also be used for training.

Unlike other attempts to do fully automated machine processing of SAR images, a top-down, knowledge based approach has been taken to fully understand what important features and information are necessary for classification. This elicitation and structuring of the knowledge is seen as a necessary step in the development of any

fully automated ice classification systems.

The techniques developed for the ice classification problem are also directly applicable to many other imaging problems requiring classification by experts.

7. ACKNOWLEDGMENT

This research was supported by the Transportation Development Centre, Transport Canada.

8. REFERENCES

1. Shaw M.L.G. and Gains B.R., A Computer Aid To Knowledge Engineering, Proceedings of the British Computer Society Conference on Expert Systems, Cambridge 1983, 263-271.
2. Kelly, G.A. The Psychology of Personal Constructs, Norton, New York, 1955.

A Fuzzy Expert System for Remote Sensing Image Analysis

Fangju Wang

Faculty of Environmental Studies
University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

Abstract

Geographical information derived from remote sensing imagery is intrinsically imprecise. The currently used techniques in remote sensing expert system researches cannot represent and process the imprecise information properly. Alternative techniques are required.

In this paper, an expert system for remote sensing land cover change detection is described. The system has been developed by using the fuzzy techniques to deal with the impreciseness. The major techniques, including fuzzy representation of geographical information, fuzzy classification of remote sensing imagery and approximate reasoning, are discussed in detail. The system algorithm and analysis results are presented.

1 Introduction

In recent years, experiments with expert systems have taken place in remote sensing to help automate the processes of digital image analysis. In building an expert system, it is critical to develop proper representation, analysis, and reasoning techniques which are suitable for the problem to be solved.

Geographical information derived from remote sensing imagery is imprecise in nature. Many relevant concepts do not have precisely defined intension and extension. "Grass land" and "soil" are such concepts. In talking about land cover, a piece of land with sparse grass can be classified into either grass land or soil. There is not a well specified criterion for distinguishing between the two cover types. The impreciseness also results from natural variations or arises through original measurements as well as data processing. In building an expert system for remote sensing image analysis, the problem of information impreciseness must be paid attention.

However, in the current remote sensing expert system researches, the techniques for information representation, analysis and reasoning are implicitly based on classical set theory and Boolean logic which are intended for precise information. Such techniques inevitably lead to information loss and, in turn, low analysis accuracy. To improve knowledge-based remote sensing image analysis, alternative techniques must be developed which can represent and process the imprecise information well.

In this paper, an expert system for land cover change detection is described. The system monitors the process of urbanization by analyzing Landsat imagery. It detects areas of change and identifies the following types of change: from vegetation to housing construction, from housing construction to urban, from vegetation to urban, and crop rotation or others. The system performs automated image analysis. No user involvement is required during the entire process of change detection.

To deal with the impreciseness, a fuzzy representation model is used to represent geographical information, a fuzzy classification algorithm has been developed to improve classification accuracy and extract more spectral information from remote sensing data, and an approximate reasoning method has been developed to make reasoning on imprecise information.

2 Fuzzy Sets

Let X be a universe of discourse, whose generic elements are denoted x . Thus, $X = \{x\}$. Membership in a classical set A of X is often viewed as a characteristic function χ_A from X to $\{0, 1\}$ such that $\chi_A(x) = 1$ if and only if $x \in A$.

A fuzzy set [6] B in X is characterized by a membership function f_B which associates with each x a real number in $[0, 1]$. $f_B(x)$ represents the "grade of membership" of x in B . The closer the value of $f_B(x)$ is to 1, the more x belongs to B .

In classical set theory, a set has precisely defined boundaries and an element either belongs to a set or not at all. No third situation is allowed. However, in the real world, there exist sets which do not have sharp boundaries. Membership of their elements cannot be simply specified as a binary "yes" or "no" criterion, and intermediate and mixture cases exist. As noted, land cover type "grass land" and "soil" do not have a sharply defined boundary between them. Fuzzy set theory provides very useful tools to deal with such sets.

However, works in remote sensing and other geographical information processing using fuzzy sets are rather scarce. Jeansoulin et al. [3] proposed to use the principles of fuzzy set theory to combine criteria for automating multitemporal segmentation. Cannon et al. [1] developed a fuzzy c-means clustering algorithm to segment a TM image. Zeno et al. [7] used fuzzy sets for contextual classification. These approaches can improve analysis to different extents. But, they do not fully bring out the potential of fuzzy sets for geographical information handling since the fuzzy techniques were merely used in limited phases of the works. If fuzzy sets can be used as consistent representation and analysis techniques throughout the entire procedure of information processing, analysis will be further improved.

3 Fuzzy Representation of Geographical Information

Geographical information is conventionally represented in *thematic maps*. A map is a set of points, lines and areas that are defined both by their location in space and by their non-spatial attributes. Currently, the linkage between the spatial entities and their non-spatial attributes are based on the membership concept of classical set theory - an entity either has an attribute or not at all.

Probably because that a primary objective of many remote sensing applications is the classification of images to prepare thematic maps and the ground truth information required for classifier training is derived from thematic maps (which may not exist physically), information representation in remote sensing image analysis basically follows the conventional method. Each pixel can only be associated with one cover class. Such a method cannot properly represent class mixture and intermediate conditions which occur in most remote sensing images.

A pixel corresponds to a cell on the ground. Quite often, such a cell contains a mixture of surface cover classes, for example grass and underlying soil. Mixture may also take place when the size of the cell is larger than the size of the features about which information is desired. Since currently only one cover class can be assigned to a pixel, information about other component classes and deviation of the assignment cannot be represented.

Different conditions may exist within a cover class. For example, vegetation may be in different conditions which are caused by such factors as plant health, age and water content. However, these conditions cannot be differentiated in a thematic map unless more classes are defined. It is clearly inaccurate to assign the same class to fresh grass and half dry grass without specifying their differences. Introducing more classes will lead to higher analysis costs and no matter how fine the classes are defined, within-class variability may exist.

Fuzzy set theory can provide a better representation for geographical information, much of which cannot be described well by a single class. In a fuzzy representation, land cover classes can be defined as fuzzy sets and pixels are set elements. Each pixel is attached with a group of membership grades to indicate the extents to which the pixel belongs to certain classes. Pixels with class mixture or in intermediate conditions can be described by membership grades. For example, if a ground cell contains two cover types "soil" and "vegetation", it may have two membership grades indicating the extents to which it is associated with the two classes.

4 Fuzzy Classification of Remote Sensing Imagery

In this section, a fuzzy classification is described which consists of *fuzzy partition of spectral space* and *fuzzy parameters*.

A ground cover class has its spectral characteristics which depend upon the interaction of electromagnetic waves with that class. In a given radiometric band, a pixel value of a remote sensing image is a function of the reflectance from component classes of the corresponding ground cell [2]. A mixed or heterogeneous pixel has its spectral characteristics which differ from those of a homogeneous pixel. Changes in condition within a given cover class also cause variation in spectral characteristics. Theoretically, the spectral information about cover components and different conditions can be utilized. However, the current classification techniques are unable to do so.

4.1 Fuzzy partition of spectral space

In remote sensing, pixel measurement vectors are often considered as points in a spectral space. Pixels with similar spectral characteristics form groups which correspond to various ground cover classes that the analyst defines. The pixel groups are referred to as *spectral classes* while the cover classes are *information classes*. Currently, image classification algorithms decide membership of each pixel in one of predefined classes.

To classify pixels into groups, the spectral space must be partitioned into regions each of which corresponds to an information class. Traditionally, the information classes are implicitly represented as classical sets. Thus partition of the spectral space is based on the principles of classical set theory. Sharp decision surfaces are precisely

defined by some decision rules. Pixels inside a region are classified into the corresponding information class. Final output is represented in a *one-pixel-one-class* image.

A serious drawback of the conventional classification is that the information about cover mixture and intermediate conditions is lost in determining pixel memberships [5]. This drawback can be largely overcome by using the *fuzzy partition of spectral space* in which partial and multiple memberships are allowed.

Formally, a fuzzy partition of spectral space is a family of fuzzy sets F_1, F_2, \dots, F_m on universe X such that

$$\forall x \in X, \\ 0 \leq f_{F_i}(x) \leq 1 \quad \text{and} \quad \sum_{i=1}^m f_{F_i}(x) = 1 \quad (1)$$

where F_1, F_2, \dots, F_m represent the spectral classes, m is number of predefined classes, x is a pixel measurement vector, and f_{F_i} is membership function of fuzzy set F_i ($1 \leq i \leq m$). $f_{F_i}(x)$ indicates the extent to which x belongs to F_i .

The fuzzy partition can be recorded in a *fuzzy partition matrix*.

$$\begin{bmatrix} f_{F_1}(x_1) & f_{F_1}(x_2) & \dots & f_{F_1}(x_n) \\ f_{F_2}(x_1) & f_{F_2}(x_2) & \dots & f_{F_2}(x_n) \\ \dots & \dots & \dots & \dots \\ f_{F_m}(x_1) & f_{F_m}(x_2) & \dots & f_{F_m}(x_n) \end{bmatrix} \quad (2)$$

where n is number of pixels, and x_i 's are pixels ($1 \leq i \leq n$). In the fuzzy partition, if m classes are defined, each pixel is attached with m membership grades.

Compared with the conventional spectral space partition, the fuzzy partition can better represent the real situation and allows more spectral information to be utilized in subsequent analysis.

4.2 Fuzzy Parameters for Image Classification

In classifier training, statistical methods are applied to training data to generate parameters for classification algorithm. The parameters play a critical part in determining pixels' membership. Low classification accuracy is largely due to the variations between statistically generated parameters and the "real" ones.

Among the factors giving rise to the variations, an important, however so far not fully realized factor lies in the knowledge representation methods. Currently, training information is represented on a basis of *one-pixel-one-class*. Once a pixel is associated with a class, it makes a full contribution in generating statistical parameters of the class. Class mixture and intermediate conditions cannot be taken into consideration. The parameters estimated in this way must vary from the "real" ones more or less. A classification based on those parameters may thus contain mistakes.

The fuzzy representation makes it possible to calculate parameters which are closer to the "real" ones. The basic idea for the methods employed in this system is that how much a pixel belongs to a class, how much it contributes to parameters of the class. The mean and covariance matrix obtained in this way are called *fuzzy mean* and *fuzzy covariance matrix* which are defined as follows.

$$\mu_c^* = \frac{\sum_{i=1}^n f_c(x_i) x_i}{\sum_{i=1}^n f_c(x_i)} \quad (3)$$

$$\Sigma_c^* = \frac{\sum_{i=1}^n f_c(x_i) (x_i - \mu_c^*)(x_i - \mu_c^*)^T}{\sum_{i=1}^n f_c(x_i)} \quad (4)$$

where n is total number of sample pixel, f_c is membership function of class c , and x_i is a sample pixel measurement vector ($1 \leq i \leq n$).

5 Approximate Reasoning

The reasoning techniques employed in the existing expert systems are mainly based on *modus ponens* which is one of the most commonly used inferencing rules in Boolean logic. The classical *modus ponens* can be represented as

$$\begin{array}{l}
 \text{If } X \text{ is } V \text{ then } Y \text{ is } G \\
 X \text{ is } V \\
 \hline
 Y \text{ is } G.
 \end{array} \quad (5)$$

The classical modus ponens is well suited for the situations that "X is V", a proposition about facts, is either true or false. However, when dealing with the imprecise geographical information in practical situations, a proposition may have an intermediate truth value. In the case that X is V' which is similar to but not exactly identical to V, one cannot reach a meaningful conclusion by applying Equation 5. For example, when condition of a rule is "land cover is soil", but the observed fact is "70 percent of the land cover is soil and the rest is vegetation", an expert system solely depending on the classical modus ponens cannot make a right reasoning. Therefore, an approximate reasoning method is needed.

The approximate reasoning technique developed in this work is an extension of modus ponens which can be stated as

$$\begin{array}{l}
 \text{If } X \text{ is } V \text{ then } Y \text{ is } G \quad CL_1 \\
 X \text{ is } V' \quad CL_2 \\
 \hline
 Y \text{ is } G \quad CL_3
 \end{array} \quad (6)$$

where CL_1 denotes certainty level of the rule, i.e., the reliability of the knowledge that "if X is V then Y is G", CL_2 denotes certainty level of the fact, i.e., the reliability that "X is V'" is true, and CL_3 is certainty level of the conclusion "Y is G". CL_3 can be calculated by applying fuzzy number multiplication to CL_1 , CL_2 and the similarity S between V and V' which are elements of a fuzzy set F:

$$CL_3 = CL_1 * CL_2 * S. \quad (7)$$

Fuzzy number multiplication * can be expressed as:

$$f_{A \cdot B}(z) = \max_{x+y=z} (f_A(x) \wedge f_B(y)) \quad (8)$$

where A and B are fuzzy numbers and \wedge denotes taking the minimum.

The similarity S between V and V' is evaluated as

$$S = 1 - |(f_F(V') - f_F(V))| \quad (9)$$

where f_F is membership function of fuzzy set F and $||$ denotes taking absolute value.

When condition part of a rule is conjunctive subconditions, and the similarities between the subconditions and facts are S_1, S_2, \dots, S_n , the total similarity S can be calculated as:

$$S = S_1 * S_2 * \dots * S_n. \quad (10)$$

6 Algorithm of the Expert System

Image differencing and post-classification comparison form the basis of the system algorithm. The former is sensitive to the areas of change and the latter can provide information for determining change types. Combining the advantages of them, the system is able to identify both areas of change and their change types. A geographical information system (GIS) is incorporated to provide ancillary information and the expert system works on the regions that the GIS has covered. In each task, the system deals with one new image. Change detection is performed on the new image referencing previously determined land cover and other information. The system data flow is illustrated in Figure 1.

Entire process of the automated change detection consists of five major steps: (1) image differencing, (2) change mask production, (3) automated fuzzy supervised classification, (4) attribute extraction, and (5) change identification:

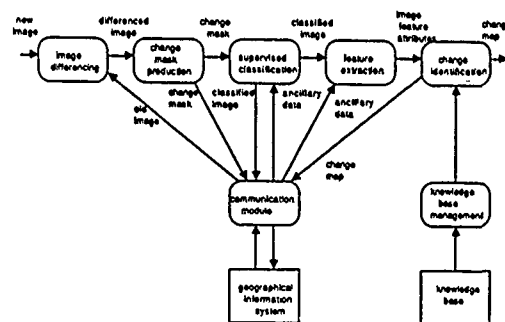


Figure 1: system data flow.

Image Differencing

In the first step, image differencing is performed on the newly input image (image of date 2) and an "old" image from which the previous land cover is determined (image of date 1). Image differencing involves the subtraction of one digital image from the other after they are precisely registered. After differencing, the areas of change can be separated in digital values. This operation is expressed mathematically as:

$$X(3)_{ijk} = X(2)_{ijk} - X(1)_{ijk} + C \quad (11)$$

where $X(1)$, $X(2)$ and $X(3)$ are pixel values at image of date 1, date 2 and the differenced image respectively, i, j , and k are line, pixel and spectral band number, and C is a constant.

Change mask production

For selecting training areas and determining potential areas of land change, a change mask is produced from the differenced image by using a K-L (Karnen-Loeve) expansion followed by a K-means clustering. The K-L expansion is an effective method for feature extraction in statistical pattern recognition. It finds an orthogonal set of basis vectors for the feature space and transforms measurement vectors into the new space. The K-means clustering algorithm classifies a digital image into K categories by minimizing the sum of squared errors within the K categories.

In the transformed differenced MSS image, the second component emphasizes brightness changes between the two original images. In this component, the land changes of interest, e.g. from vegetation to housing construction, or from housing construction to urban, are highlighted by strongly contrasted very bright or very dark tones, while the areas of no change are in intermediate tones [4]. After normalizing the pixel values, a K-means clustering is performed on the component. The very bright and dark areas are classified as potential areas of change and others as areas of no change. By assigning "1" to pixels in the potential areas of change and "0" to others, a change mask is produced.

Automated fuzzy supervised classification

An automated fuzzy supervised classification is performed on the newly input image. The change mask is used to select training areas. The previously determined land cover data from the GIS is used to train the classifier. All the areas of no change are selected as training areas.

A fuzzy set is characterized by its membership function. To perform a fuzzy partition on a spectral space, a membership function must be defined for each cover class. In this work, the membership functions are defined based on maximum likelihood classification algorithm with fuzzy mean μ and fuzzy covariance matrix Σ replacing the conventional mean and covariance matrix. The following is definition of the membership function for cover class c.

$$f_i(\mathbf{x}) = \frac{P_i^*(\mathbf{x})}{\sum_{i=1}^m P_i^*(\mathbf{x})} \quad (12)$$

where

$$P_i^*(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i^*)^T \Sigma_i^{-1} (\mathbf{x}-\mu_i^*)} \quad (13)$$

N is dimension of the pixel vectors, m is number of predefined classes, and $1 \leq i \leq m$.

Provided that the ground truth information from the GIS is sufficiently reliable, this algorithm can achieve highly accurate outputs as the biggest possible training areas are used and the fuzzy methods may generate more accurate parameters. The membership grades can provide very useful information about cover components and intermediate conditions for the subsequent analysis.

Attribute extraction

In this step, information required for change identification is collected for pixels in the potential areas of change. The information includes land cover at both dates, topographic features, allowable uses, and some historical and contextual information. The collected information is represented as attributes of the areas.

Change identification

In the last step, the areas of change and their change types are identified based on the analysis of spectral, spatial, historical and contextual attributes. Human knowledge about change identification is formulated into rules. The change identification is a reasoning process by applying the rules to the collected attributes.

The overall knowledge representation scheme employed in this system is a production system which consists of three major components: *context*, *rule base*, and *interpreter*. The factual knowledge in the context and the performance knowledge in the rule base are organized as relations.

Context of a production system is a working space for recording the initial, intermediate, and final states of a task, and performing reasoning. Context of this system also serves as a buffer for communication with the GIS.

The knowledge required for identifying land cover change are formulated into production rules which are of the form: *condition* \rightarrow *conclusion*. The condition part is a conjunction of predicate formulas. The conclusion part includes new facts for updating the context. Presently, there are about 170 rules in the rule base.

The interpreter applies the rules to the context to make reasoning. For each region selected from the change mask, the interpreter matches condition of the rules against the factual knowledge in the context to determine the change type of the region. Conditions of the rules are represented as conjunctive relation tuples and the tuples are encoded as procedures of relational operators. The procedures retrieve the context for the specified tuples. To trigger a rule is to simply activate the procedures of it. Failure of a procedure implies that one of the conditions cannot be satisfied by the facts of a given region and thus the rule cannot be fired. If all the procedures succeed, the tuple in the conclusion part is added into the context.

Approximate reasoning is conducted for identifying changes. Each change type is defined as a fuzzy set and the regions as set elements. A membership grade of a region in a change type set indicates the extent to which change of that type takes place in the region. A region may have more than one membership grades if more than one change types are involved. The membership grades are determined by using Equation 7 through 10. The final result of the change detection is a fuzzy change map in which each region is attached with one or more membership grades.

The procedural representation of rule conditions largely simplifies the interpreter in structure. It consists of only two components: *activator* and *inserter*. For each region, the activator passes param-

eters to the procedures, and activates them. The inserter adds the conclusion tuple into the context when a rule has been fired.

7 Results and Conclusion

Very encouraging results have been achieved in applying the automated change detection algorithm to the study area - Hamilton, Ontario, Canada. The fuzzy techniques have improved the analysis accuracy considerably and more information about land cover and land change has been obtained.

Fuzzy classification is performed to classify images into seven cover classes: water, industrial or commercial, residential, forest, farm land, bare soil, and pasture or other vegetation. Seven membership grades are assigned to each pixel. Tests with aerial photographs have verified that a membership grade of a pixel is proportional to the area of the corresponding cover class in the ground cell. The information about pixel component cover classes conforms very well to the real situations. Overall classification accuracy levels of about 91 percent have been measured. Compared with those from the conventional classification on the same images, an improvement of about 6 percent has been achieved [5].

By using the approximate reasoning method, the system can draw more meaningful land change information from imprecise, incomplete or unreliable information, and provide more information about the nature of change. Four membership grades are associated with each region for the four types of land change. Besides indicating the certainty levels of change identification, the membership grades may give a better description of land change. For example, maturity of a new residential area can be determined by analyzing combination of the membership grades. Change detection accuracy has been assessed on the 1974, 1978, 1981 and 1984 images of the study area. The accuracy levels for detecting the areas of change are about 95 percent, and the accuracy levels for identifying change types are about 86 percent.

References

- [1] Cannon, R.L., J.V. Dave, J.C. Bezdek, and M.M. Trivedi, 1986, "Segmentation of a Thematic Mapper Image Using the Fuzzy c-Means Clustering Algorithm", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-24, No. 3, pp. 400-408.
- [2] Fung, A.D., and F.T. Ulaby, 1983, "Matter-Energy Interaction in the Microwave Region" (Chapter 4), in R.N. Colwell ed. *Manual of Remote Sensing (Second Edition)*, Virginia: American Society of Photogrammetry.
- [3] Jeansoulin, R., Yves Fonyaine, and Werner Frei, 1981, "Multi-temporal Segmentation by Means of Fuzzy Sets" *1981 Machine Processing of Remotely Sensed Data Symposium*, pp. 336-339.
- [4] Wang, F. and R. Newkirk, 1987, "Design and Implementation of a Knowledge Based System for Remotely Sensed Change Detection", *Journal of Imaging Technology*, Vol. 13, pp. 116-122.
- [5] Wang, F., 1989, *Integrating Expert Vision Systems and Spatial Databases by Unifying Knowledge Representation Schemes: Development in Remote Sensing Image Analysis Expert Systems and Geographical Information Systems*. Ph.D. dissertation. University of Waterloo, Ontario, Canada.
- [6] Zadeh, L.A., 1965, "Fuzzy Sets", *Information and Control*, Vol. 8, pp. 338-353.
- [7] Zenzo, S.D., R. Bernstein, S.D. Degloria and H.G. Kolsky, 1987, "Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No. 6, pp. 805-814.

ESSII : EXPERT SYSTEM FOR SATELLITE IMAGE INTERPRETATION

CHITTI V S SARMA+
B.E. PRASAD+

R.KRISHNAN *
A K PUJARI +

+ Artificial Intelligence Laboratory
University of Hyderabad
Central University (PO)
HYDERABAD - 500 134
INDIA

*HEAD, SATD
National Remote Sensing Agency
(Dept. of Space, Govt. of India)
Balanagar
Hyderabad - 500 037
INDIA

ABSTRACT

Expert Systems are among the most exciting new developments in software technology. Emerging as a practical application of research in ARTIFICIAL INTELLIGENCE, these programs embody knowledge of a particular application area combined with an inference capability, which enables the programs to reach (or even exceed) the performance of human experts. The present paper deals with various aspects of design and development of one such a system, namely ESSII. The goal of the present task is to automate the interpretation process regarding geomorphic features of a satellite imagery. As a part of future plan, which is aimed at achieving complete automation of the interpretation process, the present system (ESSII) is developed using MPROLOG on APOLLO WORKSTATION. ESSII is enriched with various features such as dealing with uncertainty, explanation capability, user interface in form of different menu, and an on line help with graphical display.

KEY WORDS

Remote Sensing, Geomorphology, Expert System, Image Interpretation, Image Processing, Edge Segmentation, quad tree structure.

INTRODUCTION

Image interpretation is defined as the act of examining images for the purpose of identifying objects and judging their significance (5). The basic elements that are dealt by ESSII for the interpretation process are: tone or color, size, shape, texture, pattern.

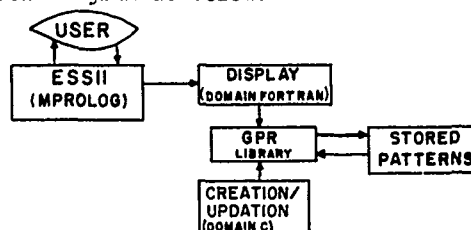
The domain around which the knowledge base is developed is the GEOMORPHOLOGY. The major landforms that are taken into consideration for the interpretation are: Hills/ Mountains, Piedmont zones, Valleys, Plains. These landforms are in turn classified into various other landforms. As the satellite

data is obtained in large volume and its interpretation is cumbersome and time consuming, and also as the precious resource of human expertise is fragile and transient one, the need of Expert System is realised. Basically, the current problem is heuristic in nature. It can not be solved mathematically or there is no algorithmic procedure to solve the same, and hence expert system approach is resorted to solve the problem.

In view of the above facts, it is evident that an Expert System can handle the problem in hand more efficiently. ESSII is one such Expert System for satellite Image Interpretation.

ENVIRONMENT AND DESIGN OVER VIEW

The aim of this project is to develop an expert system for satellite image interpretation taking the geomorphological aspects into consideration. The knowledge base is encoded in MPROLOG. The control strategy is the unification with backtracking, which is an underlying computational mechanism in MPROLOG. The system interacts with the user during a question answering session. The system can be represented in a block diagram as below:



The core of the system is the ESSII, which contains knowledge about the domain. During session with the system, this module also calls DISPLAY module. This display module is implemented using DOMAIN FORTRAN. The DISPLAY module access bitmap file containing the stored patterns. These bitmap files are created and updated by two separate modules which are implemented

in DOMAIN C. The display, create, and update modules call GPR (Graphics Primitive Resource) routines to achieve their tasks. The GPR routines are a standard piece of Domain Software.

ESSII

The knowledge about the Domain (Image Interpretation) is mainly acquired from scientists of NRSA, Hyderabad who are the authorised experts in the same. The existing literature (4, 5) also enabled us to acquire more knowledge about the domain.

After following different methodologies, we came up with a hybrid methodology and followed the same for the Knowledge Acquisition phase of development of ESSII.

Such a methodology, which is an efficient and effective blend of observational method, intuitive method and Multiple Expert approach, is found to be more practical than the individual techniques.

INCORPORATION OF UNCERTAINTY:

The conclusions drawn by the ESSII depends on two factors: 1. User input which reflects the confidence on his/her response, and 2. Factor strength that each factor is given in terms of its accounting for conclusive evidence. The way in which ESSII deals with the uncertainty is as below:

When ESSII poses a question, the user is expected to input a number in the range of +1.0 (synonymous to "yes") and -1.0 (which means "no"), it is then multiplied by factor strength to give rule confidence. Based on the rule confidence, ESSII presents its interpretation to the user.

In general though, there may be many factors, as each factor may confirm or deny a conclusion with a certain confidence

factor. In this case we followed a standard combining technique to get a single resulting confidence, that can be termed as 'current rule confidence'. Hence when all factors are exhausted, the current rule confidence will be the 'final rule confidence', based on which ESSII could come up with a conclusion and convey its confidence on the same. The combining technique, which comes from Emycin system, can be stated as below:

$$C = C1 + C2 - (C1 * C2) \text{ if } C1, C2 < 0$$

$$C = C1 + C2 + (C1 * C2) \text{ if } C1, C2 \geq 0$$

$$C = (C1 + C2) / 1 - \min(\text{abs}(C1), \text{abs}(C2))$$

Let us look at the following complex rule, which contains more than one factor.

upln:-

Write(" Let us try to identify uplands in the image"), nl

Write ("Do you see any area in the TOPOSHEET in which contour lines are closely spaced?"), nl,

reply(CF1), prod(CF1, Be-1, CF2),

write("In the FCC imagery, does this area have the following properties:

1. Tone is light to dark brown or light to dark to grey.

2. A Drainage Pattern is visible"), nl,

reply (CF3, be-1, CF4), !, combn (CF2, CF4, CF5),

write ("Do you see thin drainage lines which are joining upto form thicker lines?"), nl,

reply(CF6), prod(CF6, be-1, CF7), comb (CF5, CF7, CFB), nl,

answer (CFB, Str2), Str1="Upland

inter, msg(CFB, Str1, Str2, String1, String2, String).

In the above example, first user input is CF1, which is multiplied by 0.8(factor strength) to give CF2, which is current rule confidence (CRC). Then CF3, which is a response to the question corresponding to the second factor in the rule, multiplied by the factor strength (0.6) to give CF4. Then CF2 and CF4 are combined using the above method to give a single confidence factor (CF5), which is nothing but a CRC. Likewise the confidence factor will be propagated until all the factors are exhausted when the CRC will become Final Rule Confidence (FRC). Depending on the value of the FRC, ESSII conveys its interpretation in a most pleasing way to the user.

EXPLANATION CAPABILITY:

In ordinary life, the EXPLANATION can be a sequence of reasoning steps used to convince a person that is so. In Expert System Technology, the situation is also similar. With such a capability, ESSII can handle 'why' explanations. When ESSII asks a question, the user can answer with why, meaning "why did you ask me that question" or "what happens next on the screen". To get explanation from ESSII, the user has to choose an option of 'EXPLANATION' from the menu displayed on the screen.

USER INTERFACE:

ESSII is an interactive system and engages the user in a dialogue, where, the user is given various options at every stage to choose from. With each menu, ESSII displays messages regarding the input from the user. A system cannot be user-friendly, unless it continues to work in spite of an error in input. ESSII accepts any input and checks for its correctness and asks the user for a syntactically and semantically correct input, with the use of appropriate

user friendly messages.

The graphical interface involves the display of Drainage Patterns in which the patterns are shown in a separate window, while session with the ESSII is going on. When the user finishes his job of viewing and comparing the patterns, he/she can press 'q' or 'Q' to close the display and resume further consultation. The interface can still be made more user friendly by incorporating a Natural Language Interface.

SCOPE OF FUTURE WORK

Development of ESSII is a part of the future work, which is aimed at achieving the complete automation of satellite image interpretation. This can be realised by integrating the techniques from Image Processing, Pattern Recognition, Artificial Intelligence and conventional Computer Science. As all these are rapidly growing fields, we can find a large scope for the extension of ESSII.

Even with the partially automated interpretation, a lot of work can be done on ESSII. One major extension work can be a development of Natural Language Interface. PROLOG (with which ESSII is developed) has got special applicability to NLP(2).

Like Drainage Patterns, which are currently supported by ESSII, colour maps and texture maps can also be generated to provide the user with more information. These are also major fields where considerable research is being carried on currently.

ESSII can be made more intelligent by incorporating advanced features such as learning component, reasoning about time where data vary over time, and reasoning with time including time as an element of the reasoning process itself.

As ESSII can be used as a front end module to many other systems: Intelligent

systems for mineral exploration, ground water studies, land use studies and crop monitoring etc. can utilise the image interpretation offered by it. The modular implementation supported by MPROLOG, specially eases the improvement of Knowledge base of ESSII. We feel that it is not wise to work directly with raw imagery. The imagery should be preprocessed before actual interpretation. Preprocessing may accomplished by the following steps:

It is always advisable to work with coarser image. For this we have to divide the image into several parts. We do the major interpretation at this stage i.e. first level of interpretation of the landform features into Hills/Mountains, Piedmont zones, valleys, and plains can be done at this juncture.

The second step involves the application of segmentation techniques like edge detection and the like, by taking into consideration features like texture, tone, patterns etc. Corresponding to each such feature as may be considered, we may generate a quad tree, thus ensuring those regions with uniform feature which is under consideration are at the leaf mode of the quad tree. By merging the interpretations associated with each regions of the quad trees, we may generate a globally consistent interpretation of the entire satellite imagery. The amount of refinement that we could achieve in such an interpretation process entirely depends upon the number of quad trees that we consider.

CONCLUSIONS

It has been shown that the Satellite Image Interpretation is a valid application area for the technology of Expert System. The Knowledge Base of ESSII is encoded in MPROLOG, as it is found to be a suitable tool for the development of the same. MPROLOG's strength is not derived from a lengthy list of features, but its underlying structure based on LOGIC. The graphics interface is one of the main

features of ESSIII, that makes the system user friendly. It was achieved by the Graphics Primitive Resource library on the APOLLO Work Station. The ESSII is made powerful by incorporating the various features like Inexact Reasoning, Explanation capability, and an user interface. By the successful incorporation of Learning Component and Temporal reasoning, ESSII can be made intelligent and we can see that it is a kernel of a more substantial reasoning environment which embeds this layer of "intelligence" into the user's computational support environment.

REFERENCES

1. Shapiro. S.C., Techniques of Artificial Intelligence, Van Nostrand, New York(1979)
2. Adrian Walker, Michael Mc Cord, John F. Sowa, WG Wilson, Knowledge systems and prolog, Addison-wesely Pubs. Co.(1987).
3. Fedrick Hayes-Roth, Donald A Waterman, Doughlass B. Lenat, Building Expert Systems, Mc Braw-Hill, NY. (1983).
4. Lillesand T.M. and Kiefer R.W., Remote Sensing and Image Interpretation, John Willey, NY. (1979).
5. Manual of Remote Sensing, Vol 1 and Vol.2 American Society of Photo Grammetry(1986).
6. P.A.Subrahmanyam, "The Software Engineering of Expert System: Is Prolog Appropriate?", IEEE Transaction of Software Engineering, Vol SE 11, No.11, Nov.1985, pp 1391-1400.
7. David G.Goodenough, Morris Goldberg, Gordon Plunkett, John Zelek, "An Expert System in Remote Sensing", IEEE Trans. on Geoeengineering and Remote Sensing, Vol 25, No.3, 1987.
8. Herbert A. Simob, "Whether Software Engineering Needs to be Artificially Intelligent", IEEE Trans. on Software Engineering, Vol.SE 12, No.7, July 1986.

SHERLOCK supports the Geocoding of SAR images

M. Plößnig, W. G. Kropatsch, D. Strobl

Institute for Image Processing and Computer Graphics
Joanneum Research
Wastiangasse 6, A-8010 Graz, Austria
Telex 311265, Telefon 0316/8021-0, Telefax 0316/8021 20

Abstract

The geocoding of synthetic aperture radar (SAR) image data is performed by a transformation of a digital SAR image to a georeferenced system. A way is to search for control points in the image and a corresponding digital elevation model (DEM). In the presented approach we avoid this time-consuming search for control points by applying different feature detectors in DEM and SAR image in order to find features with the same characteristics.

This process is supported by the knowledge based system SHERLOCK which manage the various methods applicable to the special circumstances. The predominant part of SHERLOCK is the knowledge base which controls the data driven processing under the (interactive) guidance of the human operator.

Our examinations concentrate on mountainous regions where layover is dominating. An example is given to show how SHERLOCK detects layover in both the SAR image and the DEM and how it creates a consistent correspondence between the two sets of layover regions.

Keywords: expert system, knowledge base, SAR, layover, control points, geocoding

1 Introduction

In contrast to images from optical sensors, SAR images show a more complicated geometrical mapping. Most optical sensors produce line scanned images and a few of them are images with central perspective which is equivalent to human vision. SAR images, on the other hand, show a range projection. Therefore there exists a great need for transforming the image geometrically into a map projection. For this procedure - called *geocoding* - different methods exist: Domik, 1985 used image simulation; Raggam, Strobl and Triebnig, 1986 used squint angle condition and bundle adjustment; Meier and Nüsch, 1986 used doppler information and target point velocity; Kwok, Curlander and Pang, 1987 used doppler information and a three pass resampling; etc.

Most of the geocoding procedures are based on the identification of control points. These control points establish the geometrical relationship between SAR and DEM system that is used in geocoding process to resample the SAR pixels to the location of the DEM nodes. They preprocess the DEM e.g. by use of an illumination model or by simulation to create a synthetic image that bears optical resemblance with the SAR image. The coordinates of these features in the SAR image and the corresponding DEM form the set of control points used for geocoding.

The motivation for building an intelligent system that helps with the search for control points or performs the search automatically is that the search is very time consuming. The person who performs the manual search must have great skill, a lot of experience and a overall knowledge about SAR images. Even a very experienced user will find it very helpful to have an intelligent system as an aid.

SAR image and DEM are both array of numbers. The numbers in the DEM represent the elevations of the terrain and are obtained by putting a reference grid over the terrain and by using the heights in the sample grid positions. The numbers in the SAR image stand for the intensity of radar backscattering. Radar exhibits specific phenomena which distort the perceived image of the objects. One of them are layover regions which are very bright areas in radar images caused by a many to one mapping during image formation. The characteristic effect of radar layover is an increase of the grey values.

Since control points are used for geocoding the presented task of SHERLOCK is reduced to the search for control points in a SAR image and a corresponding DEM.

In order to relate SAR image and DEM for geocoding, selected points in both images are used to establish a geometrical relationship. The problem lies in the fact that points can only be identified in images by means of features like characteristic points on a layover border. Furthermore such features in the DEM and the SAR image look completely different. For example a peak is a local maximum in the DEM but not in the SAR image. To solve this problem the intelligent system needs specific knowledge about radar imagery and DEM.

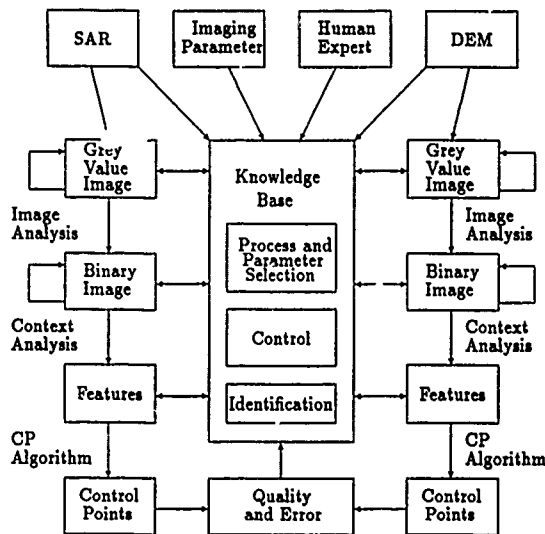


Figure 1: Search for control points with SHERLOCK

2 SHERLOCK - the Intelligent System

SHERLOCK is written in Lisp F4 on a Digital VAX 11/750 system and is an abbreviation for *a shell combining elevation models and radar images - location of control points with a knowledge base*. SHERLOCK is based on frames and rules (see Minsky, 1975). The knowledge is stored in frames. Frames are implemented as an expansion of the list structure of Lisp. They are subdivided into slots, which can in turn be subdivided into facets. Every slot or facet is allotted one or more values. They contain the knowledge of SHERLOCK as facts or as the rules which are executed to retrieve the missing information.

The search for control points is directed by an intelligent system. The choice for the next step of SHERLOCK depends on the static knowledge, the user input and the results of processes already completed. There are many methods to search for control points. They differ in speed, accuracy and failure probability. Their performance often depends on the local terrain and the image characteristics. These considerations are used to assign a priority to every method. New results may change the priorities of some processes. Interpreter facilities of Lisp make it easy to insert a new method for searching control points. The name of the method is entered in a global frame where all search methods are collected. The new method is now included in SHERLOCK. Under the influence of the knowledge in SHERLOCK the decision is made whether the new method or any of the other methods should be started or not.

Control processes are necessary to control the execution of the programs. One of those processes decides when SHERLOCK has to stop. This and other decisions are taken with the aid of user inputs. The quality of control points can be determined either by their residuals or by knowledge based methods (see 4.4).

To establish a geometrical relationship between the SAR image and the DEM it is necessary to have an even or nearly even distribution of control points (depending on the terrain). This is achieved by using a coarse grid over the DEM to decide how many control points should be found in any element of this grid.

3 Working with SHERLOCK

The knowledge base of SHERLOCK is the most predominant feature of the architecture because the knowledge constructs the basis to control the running of the program and to select the appropriate image processes. Fig. 1 gives a brief impression how the search for control points is directed by SHERLOCK.

The initial inputs are the SAR image, the DEM, the imaging parameters and the human expert knowledge. All these informations are collected in the knowledge base which is the most predominant part of the architecture because the knowledge constructs the basis to control the running of the program and to select the appropriate image processes and their parameters.

Several image processes described later in an example can be selected to modify the grey value image of the DEM and the SAR image and to obtain a binary image. The features in the binary SAR image which are qualified for a control point algorithm are selected in accordance to the binary DEM. The result of the control point algorithm are points representing a geometrical relation between the SAR image and the DEM. To complete the search a quality examination of the control points is made.

4 Searching for Control Point

Dependent upon the context of the SAR image the way is chosen to determine control points. The examinations for this paper are focused on mountainous regions where layover is dominating and therefore a bright area evaluation in the SAR image is made. Below an algorithm is proposed identifying control points based on layover areas (see M. Plößnig, 1988 and B. Billington, 1988).

4.1 Image Analysis

It is difficult to detect geographical features like peaks in a SAR image. Therefore it has been decided to concentrate in the DEM on a feature that is very distinct in SAR pictures - the layover.

Layover is not directly a feature of a DEM. Therefore a method has been used to find layover regions in the DEM. This method was suggested by Kropatsch and Strobl, 1988. It takes not only the flight path into account, but also the nadir distance, the imaging time, the slant range and the height difference of the sensor and the object point. The result is presented as a layover map.

Connectivity analysis of the binary layover map finds all connected regions which can be compared with the layover regions of the SAR image both on an individual basis and by considering different layover aggregations. The many image parts that are extracted in this way can be treated separately and in accordance to their specific properties.

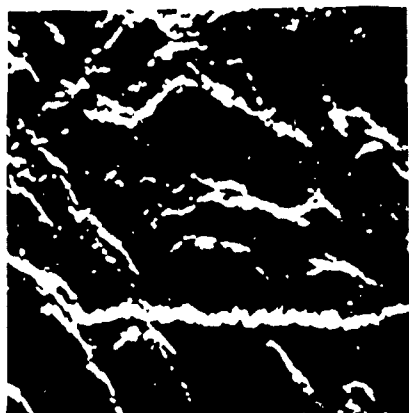


Figure 2: Layover areas of the SAR image

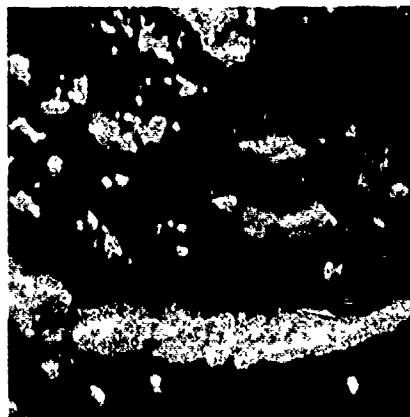


Figure 3: Layover areas of the DEM

For the examination of a *SAR* image a histogram can be used as an overview of the grey value distribution in order to select parameters for subsequent image processes. Based on the observed standard deviation an estimation of the noise can be made. This eg. is a result of an image process which will be added to the knowledge base of SHERLOCK. And in dependence of this result a filter can be chosen to smooth the *SAR* image in order to reduce the noise.

An edge-preserving filter was used to extract the layover boundaries looking to the backslopes. The intensity difference of grey values is quite high because dark areas called shadows often adjoin on these layover boundaries. To determine the boundary of layover regions closer to the sensor the covered area of layover regions in the DEM is regarded. In consideration of the resolution relation between the *SAR* image and the DEM the area of pixels representing layover can be estimated by the percentage of active layover (regions of the terrain which cause the layover) derived from the layover map of the DEM.

The lower limit is determined in such a way that the result, the binary layover image, covers the calculated percent of layover regions. The binary image contains all the pixels greater than the lower limit.

Fig. 2 and Fig. 3 show the result of layover extraction based on a *SAR* image and the corresponding DEM.

4.2 Context Analysis

Now two binary layover images are produced - one based on the *SAR* image and one based on the DEM. The features of these images are layover regions. On the basis of these features there are several ways to obtain control points. Before a control point will be selected a relation between the layover regions of the *SAR* image and the DEM has to exist. This can be done eg. by matching the layover regions of the *SAR* image and the DEM.

4.3 Control Point Algorithm

Strobl, 1988 shows that the highest position accuracy in layover regions can be found at the boundary of active layover areas. Control points of layover regions should therefore not be identified within the layover region but rather on their border. Control points on layover borders can be determined by the *minima of curvature* of the bordercurve. A description of the shapes of layover regions can be made by their minima of curvature. The similarity of a shape description of a layover border found in the *SAR* image and in the DEM allows to identify two corresponding layover regions.

4.4 Quality and Error

It is assumed that the control points are extracted at the layover borders in both the *SAR* image and the DEM. The error check for the extracted point coordinates in the DEM is done by measure the distance between the extracted point coordinates and the active layover borders in imaging direction and by measure the distance between the extracted point coordinates and the slope extremum in flight direction. Both error measurements supposed to be zero.

The a posteriori quality control of the extracted control points in the DEM is done by calculation of both single derivations of the slope gradient in flight direction and by measure the distance for the possible change of the localization on the layover border in imaging direction which is supposed to be minimum (see Strobl, 1989).

Precise conditions for the error and the quality control of the point coordinates in the *SAR* image have not been derived up to now.

| SHERLOCK | Conventional Methods |
|---------------------|----------------------|
| different processes | entire process |
| image parts | one image |
| data driven | predetermined steps |

Figure 4: SHERLOCK vs. conventional methods

4.5 Coarse-to-fine search in pyramids

In order to improve the search pyramids can be used. With the aid of a pyramid generated from a binary layover image, the most significant minima of curvature can be found in higher levels. A pyramid can be described as a stack of images of successively lower resolution (Kropatsch and Paar, 1987). They reduce the data array by a constant factor (two or four). But the accuracy decreases with the increasing levels of the pyramid. Since image processing speed depends essentially on the size of the images, major features can be detected in much less time than on the original high resolution image.

The pyramid permits analysis of coarse, but essential structures of the SAR image at a very low data volume and guides the search for the precise location of the points in the highest resolution (coarse-to-fine search). A top-down algorithm refines the position of these control points and compensates the loss of accuracy in the pyramid.

5 Results and Conclusion

SHERLOCK is a system based on frames and knowledge and controls the running of the program by its knowledge base where all informations are stored. New results of completed processes expand this knowledge base and improve the capability to select image processes and their parameters. That means the processing steps of the methods used by SHERLOCK are not predetermined like conventional methods. They are data driven.

In our examinations the search for CP's was focused on layover regions. The SAR image and the DEM - both grey value images - were transformed into binary layover images. Conventional methods used one process applied over the whole image (see Fig. 4). We extracted parts from the images - not only from the binary images but also from the grey value images - in order to apply different image processes to these image parts. This improves the efficiency because processes are not forced to be applied to the entire image.

To obtain better results with the aid of SHERLOCK more knowledge is required - not only radarspecific knowledge but also artificial intelligence techniques. Improving the knowledge base and consequently the capability to make decisions would reduce the need for a skilled user and the number of low level user interactions.

References

- [1] B. Billington. *Intelligente Paßpunktsuche im gebirgigen digitalen Höhenmodell mit SHERLOCK*. Master's thesis, Technische Universität Graz, Oktober 1988.
- [2] G. Domik. *Verfahrensentwicklung zur Analyse von digitalen Seitsicht-Radarbildern gebirgigen Geländes mittels digitaler Höhenmodelle und Bildsimulation*. PhD thesis, Technical Univ. Graz, Austria, April 1985. Also: DIBAG Report 21, Forschungsgesellschaft Joanneum, Graz, Austria.
- [3] W. G. Kropatsch and G. Paar. Aufbau einer Pyramide auf Radarbildern. In *Berichte aus Informatikinstitutionen, ÖCG-Schriftenreihe*, pages 221-232, Oldenbourg Verlag, 1987.
- [4] W. G. Kropatsch and D. Strobl. The Generation of SAR Layover and Shadow Maps from Digital Elevation Models. To be published.
- [5] R. Kwok, J. C. Curlander, and S. S. Pang. Rectification of terrain induced distortions in radar imagery. *Photogrammetric Engineering and Remote Sensing*, 53(5):507-513, May 1987.
- [6] E. Meier and D. Nüesch. Geometrische Entzerrung von Bildern orbitgestützter SAR-Systeme. *Bildmessung und Luftbildwesen*, 54(5):205-216, 1986.
- [7] M. Minsky. A Framework for Representing Knowledge. *The Knowledge of Computer Vision*, 211-277, 1975.
- [8] M. Plößnig. *Intelligente Paßpunktsuche in gebirgigen Regionen eines SAR Bildes*. Master's thesis, Technical Univ. Graz, Austria, Wastiangasse 6, Graz, October 1988. Also: DIBAG Report 39, Forschungsgesellschaft Joanneum, Graz, Austria.
- [9] J. Raggam, D. Strobl, and G. Triebnig. *The Rectification of SAR Image Data using a Digital Elevation Model and Image Simulation Techniques. Phase-B Study for ERS-1 Processing and Archiving Facility*. ESA Contract Report 6292/85/HGE-I, Technical Note 17, Forschungsgesellschaft Joanneum, Graz, Austria, August 1986.
- [10] D. Strobl. *The Effects of Control Point Location Errors on SAR Geocoding in Mountainous Terrain*. Master's thesis, Technical Univ. Graz, Austria, January 1989. Also: DIBAG Report 37, Forschungsgesellschaft Joanneum, Graz, Austria.

ATMOSPHERIC CORRECTION OF HIGH RESOLUTION LAND SURFACE IMAGES

D. J. Diner, J. V. Martonchik, E. D. Danielson, and C. J. Bruegge

Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California USA 91109

ABSTRACT

Algorithms to correct for atmospheric scattering effects in high spatial resolution land surface images require the ability to perform rapid and accurate computations of the top-of-atmosphere diffuse radiance field for arbitrarily general surface reflectance distributions (which may be both heterogeneous and non-lambertian) and atmospheric models. We have been developing such algorithms, using 3-dimensional radiative transfer (3DRT) theory. We describe the methodology used to perform the 3DRT calculations, demonstrate how these calculations are used to perform atmospheric corrections, and illustrate the sensitivity of the retrieved surface reflectances to atmospheric structural parameters.

Keywords: Atmospheric Corrections, Radiative Transfer

1. INTRODUCTION

A determination of true surface reflectance using high resolution (tens to hundreds of meters/pixel), radiometrically calibrated satellite land images must correct for the atmospheric effects of (1) extinction due to molecular and aerosol scattering and absorption, (2) path radiance introduced by scattering from the atmosphere, and (3) horizontal diffusion of radiation leaving the surface (the adjacency effect). The latter effect is particularly important for scenes containing reflectance contrasts, e.g., near coastal boundaries or inland bodies of water. Part of the process of retrieving surface reflectance includes computing top-of-atmosphere (TOA) diffuse radiance for an arbitrary atmospheric model overlaying an arbitrary surface reflectance distribution. If the model atmosphere is defined to be horizontally homogeneous, the atmospheric extinction and the path radiance can be computed using a standard 1-dimensional radiative transfer (1DRT) algorithm. The determination of the radiance arising from photons diffusely transmitted from the surface to space, however, must be computed using a 3-dimensional radiative transfer (3DRT) algorithm and is generally the most time consuming element of the atmospheric correction scheme.

In this paper we describe our progress in generating fast algorithms for computing the diffuse radiance and also in retrieving surface reflectance. Fast algorithms are a necessity if image correction for 3-D atmospheric effects is to be performed on a routine basis. This will be especially true in the Earth Observing System (Eos) era of remote sensing when vast quantities of image data will be obtained daily and must be analyzed in a timely fashion. Our current 3DRT algorithms are now comparable in speed to standard 1-D diffuse radiance algorithms, with

negligible loss of accuracy compared to our earlier versions. Our ultimate goal is a rapid algorithm that can be applied routinely to data acquired by the High- and Moderate-Resolution Imaging Spectrometers (HIRIS and MODIS) on Eos-A, and the Multi-angle Imaging SpectroRadiometer (MISR), an instrument investigation also selected for Eos-A (Diner *et al.*, 1989).

We also demonstrate the sensitivity of retrieved reflectances to uncertainties in knowledge of atmospheric parameters and compare retrieved surface reflectances using both 1- and 3-D techniques. These results illustrate the relative importance of 3DRT effects and also the ease with which the 3-D retrieval procedure can be applied in atmospheric correction of images.

2. OPTICAL TRANSFER FUNCTION

The 3-D diffuse radiance field can be represented as the convolution of an atmospheric point-spread, or blurring, function with the surface illumination distribution. We can then write the expression for the TOA radiance as (Diner and Martonchik, 1985):

$$\begin{aligned}
 I(x,y;\mu,\mu_0,\phi-\phi_0) = & \\
 & I_0(\mu,\mu_0,\phi-\phi_0) + \exp(-\tau/\mu) \times \\
 & \times \pi^{-1} \int_0^1 \int_0^{2\pi} \rho(x,y;\mu,\mu',\phi-\phi') D(\mu',\mu_0,\phi'-\phi_0) \mu' d\mu' d\phi' + \\
 & + R(x,y;\mu,\mu_0,\phi-\phi_0) \quad \{1\}
 \end{aligned}$$

where I is the TOA radiance as a function of surface spatial coordinates x and y , cosines of the view and illumination angles μ and μ_0 , and the relative solar azimuth angle $\phi-\phi_0$. The first term on the right-hand-side of Eq. 1, the atmospheric path radiance I_0 , is the radiance reflected by the atmosphere without any surface interaction; the second term is the direct radiance field, the field which is not scattered or absorbed by the atmosphere upon leaving the surface; and the third term, R , is the diffuse radiance field. The atmosphere-related quantities in Eq. 1 are functions of the opacity τ , single-scattering albedo w , vertical distribution of the scatterers, and the scattering phase function.

The diffuse radiance field can be expressed as

$$R(x,y;\mu,\mu_0,\phi-\phi_0) = \pi^{-1} \int_0^1 \int_0^{2\pi} T(x,y;\mu,\mu',\phi-\phi') \otimes S(x,y;\mu',\mu_0,\phi'-\phi_0) d\mu' d\phi' \quad \{2\}$$

where S is the reflected surface illumination distribution, T is the atmospheric point-spread function (or upward diffuse transmittance), and the symbol \otimes denotes a convolution over x and y . The reflected surface illumination distribution, in turn, is defined as

$$S(x,y;\mu,\mu_0,\phi-\phi_0) = \int_0^1 \int_0^{2\pi} \rho(x,y;\mu'',\mu_0,\phi''-\phi_0) D(\mu'',\mu_0,\phi''-\phi_0) d\mu'' d\phi'' \quad \{3\}$$

where D is the total downward directed radiance at the surface and ρ is the spatially-varying surface directional reflectance distribution. Multiple reflections between the atmosphere and surface give rise to some dependence of D on the average value of ρ .

The characteristic width of the point-spread function (PSF) is on the order of the effective scale height of the scatterers in the atmosphere (Diner and Martonchik, 1985). The atmosphere therefore behaves as a low-pass spatial frequency filter and, as such, the computation of the diffuse radiance field may be performed more efficiently in the spatial Fourier transform domain. The Fourier analog of the atmospheric PSF is the optical transfer function (OTF), which is generally complex, except for nadir viewing, whence the OTF is real and equivalent to its modulus, known as the modulation transfer function (MTF). In this special case both the PSF and OTF are circularly symmetric.

If the atmosphere is horizontally homogeneous, both the path radiance and the direct radiance field can be computed using a standard 1DRT algorithm since the downward directed radiance, D , has, to a good approximation, no horizontal variability. The diffuse radiance field, as noted above, is best calculated in the Fourier domain. Therefore, taking the Fourier transform of Eq. 2, we find

$$\bar{R}(u,v;\mu,\mu_0,\phi-\phi_0) = \pi^{-1} \int_0^1 \int_0^{2\pi} \bar{T}(u,v;\mu,\mu',\phi-\phi') \bar{S}(u,v;\mu',\mu_0,\phi'-\phi_0) d\mu' d\phi' \quad \{4\}$$

where the overbars indicate the Fourier transform, u and v are the spatial frequencies corresponding to x and y , and the convolution operation has transformed into multiplication. Note that the OTF at zero spatial frequency, $T(0,0;\mu,\mu',\phi-\phi')$, is just the 1-D upward diffuse transmittance and is calculated using the same 1DRT code used to calculate path radiance. As such, all orders of scattering are automatically computed. For spatial frequencies not equal to zero, however, the OTF is more efficiently computed using the method of successive orders of scattering, where we find that truncating the computations after two atmospheric scattering events is a sufficiently accurate representation of the total OTF.

The shape of the OTF is sensitive to a number of atmospheric structural parameters, including the opacity and scale height of molecular and aerosol scatterers, and the asymmetry

parameter of the aerosol phase function. This sensitivity to the atmosphere plus the large number of OTF values needed, corresponding to the different spatial wavenumbers, makes the OTF computation algorithm the largest consumer of time in the diffuse radiance calculation. An investigation of ways to appreciably reduce the OTF computing time without significantly affecting accuracy resulted in the following measures: (1) treating the double scattering events as isotropic scattering, (2) creating look-up tables for the various functions particular to the OTF calculation (Diner and Martonchik, 1984a, 1984b), and (3) interpolating the OTF in spatial frequency using empirical functions fitted to values calculated at only a few wavenumbers. These measures, together, resulted in two orders of magnitude increase in the computational efficiency of the OTF. The OTF for the atmospheric model described in the next section is shown in Fig. 1. The viewing angle is nadir so that the function is circularly symmetric in spatial frequency, and the incident angles have been integrated over the hemisphere in correspondence with a lambertian surface reflectance.

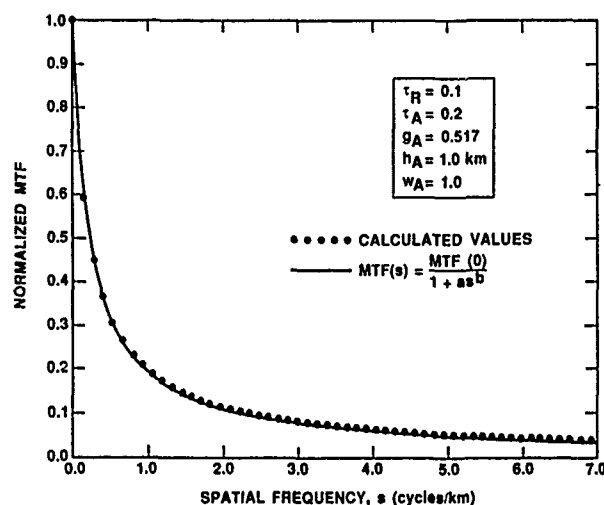


Figure 1. Optical transfer function calculated at the discrete spatial frequencies (dots) and using an empirical three-point interpolating function (solid line)

3. SURFACE REFLECTANCE RETRIEVAL

The simulated scene studied in this paper was generated from part of a radiometrically calibrated Landsat Thematic Mapper image of the coastal region near Los Angeles. The radiance values of the selected portion (256 x 256 pixels at 30-m resolution) of the TM image were transformed to effective lambertian albedo and the darkest pixel value then subtracted from the whole scene, resulting in a simulated no-atmosphere surface scene in which the darkest pixel is black. An arbitrary, normalized reflectance distribution can be introduced at this point to create a more realistic surface. However, in order to see, unambiguously, the magnitude of the adjacency effect in the following surface reflectance retrievals, the reflectance distribution for the scene studied in this paper was taken to be lambertian for all pixels in the image. The model atmosphere used to calculate simulated TOA radiances was horizontally homogeneous with two scattering components: (1) Rayleigh scattering molecules with an opacity, τ_R , of 0.1, a scale height, h_R , of 8 km and a single-scattering albedo, w_R , of unity and (2) a Mie scattering aerosol with an opacity, τ_A , of 0.2, a scale height, h_A , of 1 km, a

single-scattering albedo, w_A , of unity and a phase function asymmetry parameter, g_A , of 0.517. The wavelength was assumed to be 550 nm. The viewing geometry was nadir with the sun angle 38° off of nadir.

Because the surface reflectance distribution is lambertian, Eq. 4 can be written as

$$\bar{R}(u,v;\mu,\mu_0,\phi-\phi_0) = \bar{\rho}(u,v) \bar{T}(u,v;\mu,\phi) D(\mu_0,\phi_0) \quad \{5\}$$

where

$$\bar{T}(u,v;\mu,\phi) = \int_0^1 \int_0^{2\pi} \bar{T}(u,v;\mu',\phi-\phi') d\mu' d\phi' \quad \{6\}$$

and

$$D(\mu_0,\phi_0) = \pi^{-1} \int_0^1 \int_0^{2\pi} D(\mu'',\mu_0,\phi''-\phi_0) \mu'' d\mu'' d\phi'' \quad \{7\}$$

Taking the Fourier transform of Eq. 1, we then have

$$\begin{aligned} \bar{I}(u,v;\mu,\mu_0,\phi-\phi_0) = \\ I_0(\mu,\mu_0,\phi-\phi_0) \delta_{u0} \delta_{v0} + \exp(-\tau/\mu) \bar{\rho}(u,v) D(\mu_0,\phi_0) \\ + \bar{T}(u,v;\mu,\phi) \bar{\rho}(u,v) D(\mu_0,\phi_0) \end{aligned} \quad \{8\}$$

where δ_{ij} is a Kronecker delta. This expression can immediately be recast as an algorithm for surface reflectance retrieval:

Step I. Fourier transform the image to be corrected from the spatial domain to the frequency (wavenumber) domain.

Step II. Compute $\bar{\rho}(u,v)$,

$$\bar{\rho}(u,v) = \frac{[\bar{I}(u,v;\mu,\mu_0,\phi-\phi_0) - I_0(\mu,\mu_0,\phi-\phi_0) \delta_{u0} \delta_{v0}]}{[\exp(-\tau/\mu) + \bar{T}(u,v;\mu,\phi)] D(\mu_0,\phi_0)} \quad \{9\}$$

where the path radiance I_0 , the direct transmittance $\exp(-\tau/\mu)$, the OTF \bar{T} , and the downward directed radiance D are calculated from *a priori* knowledge of the basic atmospheric structural parameters. Since D has a slight dependence on the average albedo, $\bar{\rho}(0,0)$, Eq. 9 can be iterated for the case $u,v = 0$ to obtain both D and $\bar{\rho}(0,0)$.

Step III. Fourier transform $\bar{\rho}(u,v)$ back to $\rho(x,y)$.

Since Eq. 9 includes both measurements [the Fourier transform of the radiance image $I(x,y;\mu,\mu_0,\phi-\phi_0)$] and quantities calculated from supposedly known atmospheric quantities, any uncertainty in the atmospheric quantities will result in corresponding uncertainty in the retrieved surface reflectance $\rho(x,y)$. Using the simulated Landsat-based image described above, we show in Figs. 2 - 5 retrieved surface reflectance sensitivity results to each of four basic aerosol parameters: scale height, single-scattering albedo, phase function asymmetry factor, and optical depth.

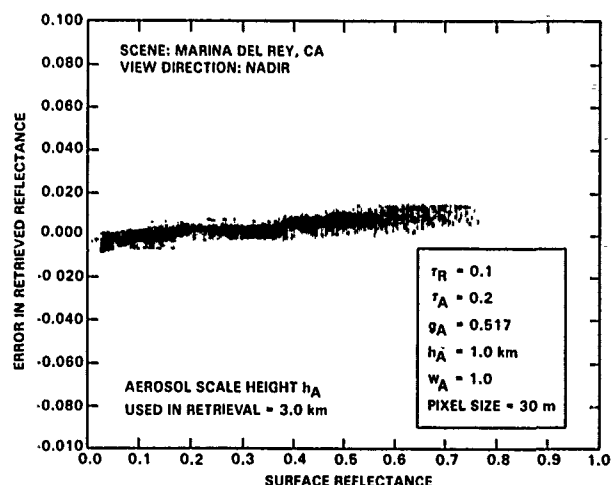


Figure 2. Surface reflectance retrieval with error in aerosol scale height

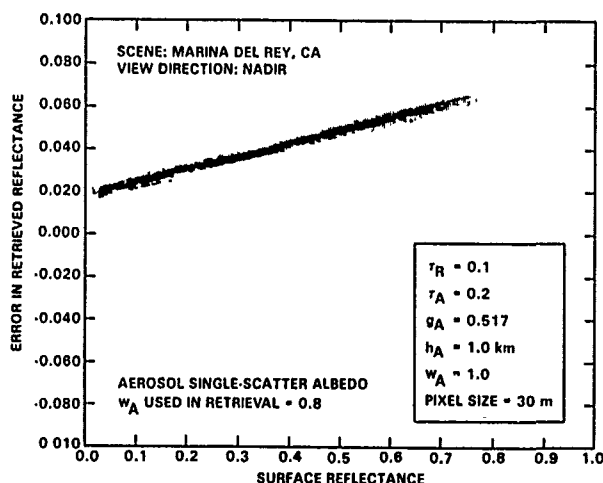


Figure 3. Surface reflectance retrieval with error in aerosol single-scattering albedo

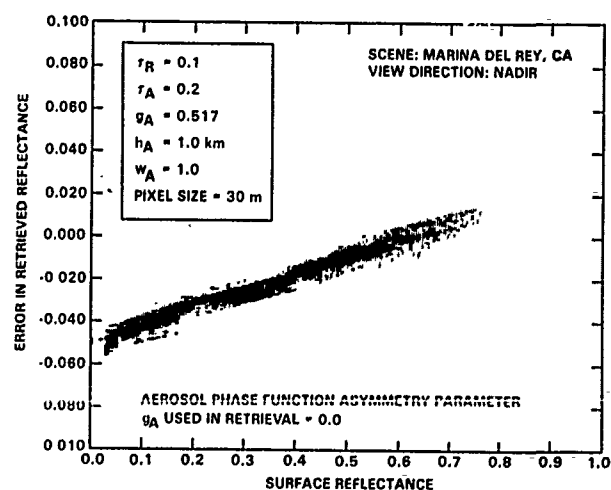


Figure 4. Surface reflectance retrieval with error in aerosol phase function

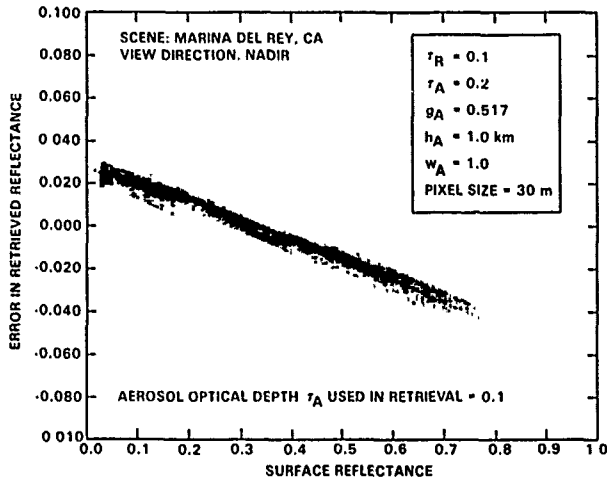


Figure 5. Surface reflectance retrieval with error in aerosol optical depth

These results show that significant errors in surface reflectance can occur if the atmospheric conditions are not determined well enough. However, a potentially more important source of error is ignoring the 3DRT effects and using a 1-D retrieval scheme. From Eqs. 1, 6, and 7 we can write the 1-D retrieval algorithm as

$$\rho(x,y) = \frac{[I(x,y;\mu,\mu_0,\phi-\phi_0) - I_0(\mu,\mu_0,\phi-\phi_0)]}{[\exp(-\tau\mu) + \bar{T}(0,0;\mu,\phi)] D(\mu_0,\phi_0)} \quad (10)$$

where $\bar{T}(0,0;\mu,\phi)$ is equivalent to the 1-D upward diffuse transmittance.

Figure 6 shows the retrieved surface reflectance of the simulated scene using Eq. 10.

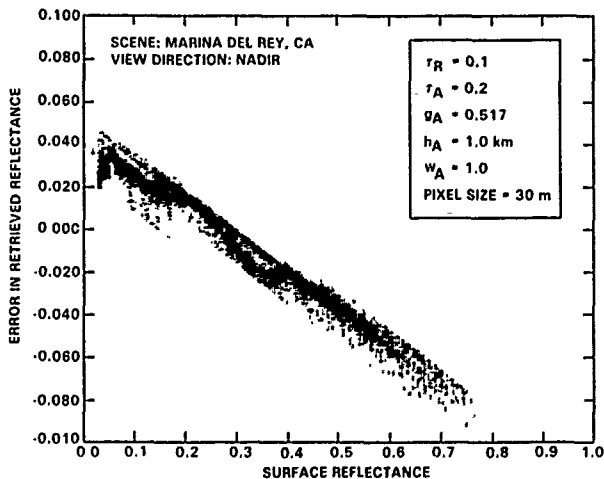


Figure 6. Surface reflectance retrieval using a 1-dimensional retrieval scheme

Comparing the results of Fig. 6 with those of Figs. 2 - 5 shows that ignoring the adjacency effect by using a 1-D retrieval can result in significantly more error in the retrieved surface reflectance than imprecise knowledge of the structure of the atmosphere. This is especially true in for those regions of a scene where albedo contrast is large.

4. CONCLUSIONS

Retrieval of surface reflectance, taking into account 3-D radiative transfer effects, requires the computation of the optical transfer function (or, alternatively, the point-spread function) of the atmosphere. Considerable progress has been made in simplifying the OTF computational process, allowing a two orders of magnitude increase in speed and making that part of the retrieval algorithm timewise comparable to the necessary 1-D computations. We expect additional improvements in speed in both the 1- and 3-D aspects of the algorithm in the future. Using Landsat-based, simulated satellite imagery, we have demonstrated the sensitivity of retrieved surface reflectance to basic atmospheric structural parameters and also compared results using both a 1-D and a 3-D retrieval approach. For scenes with high albedo contrast we find that using a 1-D algorithm and having perfect knowledge of the atmosphere can introduce more error into the surface reflectance retrieval than using a 3-D algorithm and knowing the atmosphere to a far less perfect but more reasonable degree.

5. REFERENCES

1. Diner, D. J., Bruegge, C. J., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Sellers, P. J., Clark, J., Daniels, J. A., Danielson, E. D., Duval, V. G., Klaasen, K. P., Lilienthal, G. W., Nakamoto, D. I., Pagano, R. J., and Reilly, T. H., "MISR: A Multiangle Imaging SpectroRadiometer for geophysical and climatological research from Eos", *IEEE Trans. Geosci. Rem. Sensing*, Vol. GE-27, p. 200, 1989.
2. Diner D. J. and Martonchik J. V., "Influence of aerosol scattering on atmospheric blurring of surface features", *IEEE Trans. Geosci. Rem. Sensing*, Vol. GE-23, p. 618, 1985.
3. Diner D. J. and Martonchik J. V., "Atmospheric transfer of radiation above an inhomogeneous non-lambertian reflective ground. I. Theory", *JQSRT*, Vol. 31, p. 97, 1984a.
4. Diner D. J. and Martonchik J. V., "Atmospheric transfer of radiation above an inhomogeneous non-lambertian reflective ground. II. Computational considerations and results", *JQSRT*, Vol. 32, p. 279, 1984b.

This research was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

SURFACE REFLECTANCE RETRIEVAL USING ATMOSPHERIC CORRECTION ALGORITHMS

P.M. Teillet

Canada Centre for Remote Sensing
1547 Merivale Road, Nepean, Ontario, Canada K2G 3J4

ABSTRACT

Atmospheric correction algorithms are discussed in the context of surface reflectance retrieval given data acquired by satellite or aircraft sensors. It is shown that the use of an inappropriate method in such a computation can give rise to substantial errors in surface reflectance prediction. The use of a monochromatic atmospheric calculation to characterize a spectral band result can also generate significant errors in retrieved surface reflectance in commonly used sensor bands.

Keywords: reflectance, atmospheric correction, spectral bands.

1. INTRODUCTION

A primary use of atmospheric radiative transfer codes in remote sensing is the determination of surface reflectance values given image data acquired by satellite or aircraft sensors. Section 2 of this paper outlines two different approaches to surface reflectance predictions and the problems that can arise if an inappropriate method is used. In Section 3, the question of monochromatic versus bandpass calculations is addressed by means of simulation studies.

2. SURFACE REFLECTANCE RETRIEVAL

Atmospheric codes are typically formulated to calculate the apparent radiance or reflectance at the sensor using the surface reflectance as an input boundary condition. This may be referred to as a 'forward' run of the code. Algorithms that use numerical methods to solve the equation of radiative transfer can be changed to work in the opposite direction by means of iterative or mathematical inversion approaches. Analytical expressions can usually be solved algebraically for the unknown surface reflectance and thus be rewritten in a form suitable for computation directly in the 'reverse' mode. An example of the latter type of code is the 5S code from France (Tanré *et al.*, 1966), which makes extensive use of analytical expressions and preselected atmospheric models, resulting in a very short execution time. The expression '5S' stands for simulation of the satellite signal in the solar spectrum. Despite its approximate nature, the 5S code has proven to be significantly more accurate than other algorithms with fast or even moderate execution times (Royer *et al.*, 1988). Consequently, it is of interest to modify the 5S code, which is approximate but still reasonably accurate, for surface reflectance retrieval

computations (Teillet *et al.*, 1989).

The 5S code takes into account Rayleigh and aerosol scattering, as well as gas absorption due to water vapour, carbon dioxide, ozone, and oxygen. Input parameters can be chosen from preprogrammed standard conditions and models or specified by the user. The main outputs of the program are the various components of the apparent reflectance (and radiance) at satellite altitude, the different components of the irradiance at the surface, and the scattering and gaseous transmittances. For a surface of uniform reflectance and in the formulation of Tanré *et al.* (1986),

$$\rho^*(\theta_s, \phi_s, \theta_v, \phi_v, \lambda) = \quad (1)$$

$$r_g(\theta_s, \theta_v, \lambda) \left[\frac{\tau(\theta_s, \lambda) \tau(\theta_v, \lambda) \rho(\lambda)}{1 - \rho(\lambda) s(\lambda)} + \rho_a(\theta_s, \phi_s, \theta_v, \phi_v, \lambda) \right],$$

where

- ρ^* = apparent spectral reflectance at the sensor, or spectral planetary albedo,
- τ_g = gaseous transmittance,
- $\tau(\theta_s)$ = $\exp(-\delta/\mu_s) + \tau_d(\theta_s)$
= downward scattering transmittance,
- $\tau(\theta_v)$ = $\exp(-\delta/\mu_v) + \tau_u(\theta_v)$
= upward scattering transmittance,
- τ_d = diffuse transmittance,
- θ_s = solar zenith angle,
- θ_v = sensor zenith angle,
- μ_s = $\cos \theta_s$,
- μ_v = $\cos \theta_v$,
- ϕ_s = solar azimuth angle,
- ϕ_v = sensor azimuth angle,
- δ = optical thickness,
- s = spherical albedo of the atmosphere,
- ρ_a = atmospheric reflectance,
- ρ = surface reflectance,
- λ = wavelength.

For a given spectral band and surface reflectance, $\rho(\lambda)$, the 5S atmospheric code run in the forward direction computes $\rho^*(\lambda)$ at 0.005-micrometer intervals across the band, and then integrates these values to give the band-integrated apparent reflectance at the sensor, ρ^* , according to the equation

$$\rho^* = \frac{\int \rho^*(\lambda) B(\lambda) E_0(\lambda) d\lambda}{\int B(\lambda) E_0(\lambda) d\lambda} \quad (2)$$

where $B(\lambda)$ is the relative spectral response profile of

the sensor bandpass and $E_0(\lambda)$ is the exo-atmospheric solar irradiance. Band-integrated values are also calculated in the same manner for the other quantities in equation (1) and provided as outputs of the 5S code run.

If the surface reflectance is to be determined, the code has to be run in the reverse direction. Solving equation (1) for $\rho(\lambda)$ yields the non-linear expression

$$\rho(\lambda) = \frac{y(\lambda)}{[1 + y(\lambda)s(\lambda)]} \quad (3)$$

$$\text{where } y(\lambda) = \frac{\rho^*(\lambda)}{\tau_g(\lambda)\tau_t(\lambda)} - \frac{\rho_a(\lambda)}{\tau_t(\lambda)} \quad (4)$$

$$\tau_t(\lambda) = \tau(\theta_s, \lambda) \tau(\theta_v, \lambda) \quad (5)$$

and angular dependencies have generally been omitted for convenience. Because it is assumed to originate from a given satellite sensor band, the apparent reflectance, $\rho^*(\lambda)$, is a constant equal to ρ^* , which is an input to the reverse 5S run.

The non-linearity in $\rho(\lambda)$ in equation (1) and the consequent non-linearity in the reverse equation (3) are due to the multiple scatterings of radiation between the surface and the atmosphere, i.e., the factor of $[1 - \rho(\lambda)s(\lambda)]$ in the denominator of equation (1). Many atmospheric codes do not take this effect into account. A test case was run using 5S in order to illustrate the magnitude of this effect if it is ignored when predicting surface reflectances. The example was for a vegetated surface and atmospheric conditions consisting of a mid-latitude summer model with continental aerosols and a visibility of 23 kilometres. The solar zenith angle was taken to be 25° and nadir observation geometry was assumed. Results for a wavelength of 0.480 micrometers are given in Figure 1. Although most vegetated targets of interest in remote sensing have relatively low reflectances in this spectral region, errors of several percent can still arise if multiple scattering is ignored. At 0.775 micrometers, the errors are roughly one-third of those shown in Figure 1. However, vegetation reflectances are considerably higher than in the blue and so the errors will still be on the order of several percent.

The reverse computation can be implemented in two different ways. One approach (Method A) is to use the band-integrated values for τ_g , τ_t , and ρ_a to form the expression

$$\rho = y/(1+ys) \quad (6)$$

$$\text{where } y = \frac{\rho^*}{\tau_g \tau_t} - \frac{\rho_a}{\tau_t} \quad (7)$$

Recall that the apparent reflectance ρ^* is the input, assumed to have been obtained from the satellite sensor observation in that band. The other approach (Method B) is to evaluate equations (3) and (4) for $\rho(\lambda)$ on a grid of wavelengths across the band and integrate as follows:

$$\rho = \frac{\int \rho(\lambda) B(\lambda) E_0(\lambda) d\lambda}{\int B(\lambda) E_0(\lambda) d\lambda} \quad (8)$$

In evaluating equation (4), the quantities $\tau_g(\lambda)$, $\tau_t(\lambda)$, and $\rho_a(\lambda)$ are computed by the code, whereas $\rho^*(\lambda)$ is a constant equal to the input ρ^* . Method A is more practical because slightly less computation time is

involved and, more importantly, the apparent reflectance at the sensor ρ^* is available only as a band-integrated value anyway. However, Method B appears to be a more proper approach at first glance.

The two approaches to the reverse computations do not give the same answer and it has proven to be instructive to analyze how this comes about. Consider the test case summarized in Table 1. Briefly, the spectral bands are assumed to correspond to the reflective bandpasses of the Landsat-5 Thematic Mapper (TM), SPOT-1 High Resolution Visible (HRV) and NOAA-9 Advanced Very High Resolution Radiometer (AVHRR) sensors, viewing the earth at an off-nadir angle of 30 degrees (TM is obviously a hypothetical case in this respect). The illumination and atmospheric conditions are as indicated in the table and the uniform surface reflectance is given by a standard vegetation spectrum.

In order to test the reverse runs of 5S, a forward run was executed for each spectral band, thereby generating apparent reflectance values for use as input to the reverse runs. The surface reflectances predicted by the reverse runs can then be compared to the original surface reflectance used as input to the forward case. The input conditions listed in Table 1 were used to initiate this sequence, and so the surface reflectance for vegetation, $\rho(\lambda)$, is a spectral distribution as a function of wavelength. However, the 5S code also computes a band-integrated value (using equation (8)) to report to the user. This is the reference value used in the comparison since the outputs of the reverse runs are a single value for a given spectral band.

Results are presented in Table 2. The predicted surface reflectance based on Method A is denoted ρ_A and that based on Method B is called ρ_B . As indicated in the table, ρ_A (obtained from band-integrated atmospheric values) is within one percent of the original surface reflectance in all bands. The results for ρ_B (itself band-integrated) are not as close and there are evidently substantial differences in TM band 5 and AVHRR band 2. The problem arises because the atmosphere adds a wavelength dependence to the surface vegetation spectrum that the sensor cannot capture explicitly and that is not taken into account at the input to the reverse calculation. Because the apparent reflectance, ρ^* , is assumed to be based on satellite sensor data in a given spectral band, $\rho^*(\lambda)$ as it is input to the Method B reverse run is constant across the band and equal to the band-integrated value, ρ^* . However, the actual spectral distribution, $\rho^*(\lambda)$, deviates quite markedly from the band-integrated value, ρ^* , in the longer wavelength portion of TM band 5, for example (Figure 2). The wavelength-specific nature of the Method B computation then over-corrects this constant value for strong atmospheric effects in that part of the bandpass (particularly gas absorption effects). Note that it is because this deviation is asymmetrical in the bandpass that a problem arises. In other bands where there is significant gaseous absorption, the net deviation can be small after integration.

3. BANDPASS VERSUS MONOCHROMATIC COMPUTATIONS

The more accurate models of radiative transfer through the atmosphere are complex and can require a lot of computer processing time. In remote sensing studies, such models are typically run at just one wavelength to represent a spectral band in order to save time. A wavelength often chosen for this purpose is the equivalent centre wavelength for the bandpass based on a moments method (Palmer and Tomasko, 1980). However, the question arises as to whether a monochromatic result

is sufficient for wider bandwidths. More approximate atmospheric codes run faster and can therefore be used at many wavelengths across a given spectral band. Such codes are also more easily incorporated into the image correction framework, where the atmospheric state must be assessed at a number of locations in the scene (Teillet *et al.*, 1987).

Because the 5S code has a relatively short execution time, it computes atmospheric parameters on a 0.005-micrometer grid and integrates over the chosen spectral band. This approach is less feasible with detailed numerical codes that take a long time to run and so a monochromatic value is often used to characterize a spectral band result. For wider bandwidths such as NOAA AVHRR channel 1 (0.530-0.810 micrometers) and channel 2 (0.680-1.170 micrometers), the single-wavelength calculation may not be very representative. The 5S code was used to explore this issue without using excessive amounts of computer time.

In one comparison, the apparent reflectances obtained from forward 5S runs based on Table 1 were used as inputs to two reverse run cases. One case consisted of full bandpass calculations to predict surface reflectances in the TM, HRV, and AVHRR reflective bands, where a full bandpass calculation consists of the aforementioned integration by the 5S code using a 0.005-micrometer grid. In the other case, an equivalent band centre wavelength was obtained using the moments method for a given band and a 5S reverse run was performed monochromatically at that wavelength. The results are presented in Table 3. The differences between predicted surface reflectances based on bandpass versus monochromatic computations are at least 2% and can be as high as 22% (AVHRR channel 2). The difference between the two cases in any given situation will depend on the spectral dependence of the exo-atmospheric solar irradiance, the surface reflectance, the atmospheric characteristics, and the sensor bandpass response.

In another comparison, both bandpass and monochromatic modes were used to generate apparent radiance values at sensor altitude in NOAA-9 AVHRR channel 2. Results were also obtained from monochromatic runs at 4 wavelengths in the bandpass, as well as at 12 wavelengths. The spectral radiances obtained for the various component wavelengths of the band were then combined by means of weighting coefficients based on sub-band response values normalized by the total filter response for that case. These cases for limited numbers of wavelengths were compared to the full bandpass integration which, being based on a 0.005-micrometer grid, involved dozens of wavelengths across the spectral band. The input conditions for the 5S code runs are given in Table 4 and involve uniform sand and uniform vegetation surfaces. The results given in Figure 3 indicate that single-wavelength radiances differ greatly from the bandpass value. In AVHRR channel 2, this occurs primarily because one wavelength cannot represent the gas transmittance variations and the shape of the skewed sensor response profile across the band, regardless of the type of surface reflectance. However, the convergence toward the reference bandpass result with an increasing number of wavelengths appears to be fairly rapid. A comparable analysis for AVHRR channel 1 resulted in a slower convergence toward the reference case for a vegetation target.

4. CONCLUDING REMARKS

A reverse computation capability for surface reflectance retrieval has been implemented in the 5S atmospheric code in two different ways. The two

approaches do not always give the same answer. The problem has been studied with the help of simulations involving the reflective spectral bands of the Landsat TM, SPOT HRV, and NOAA AVHRR sensors, observing a vegetated surface through a standard atmospheric state. For sensor bands where there is significant gaseous absorption distributed asymmetrically across the band, the spectral distribution of apparent reflectance at sensor altitude can deviate quite markedly at certain wavelengths from the band-integrated value observed by the sensor. In such cases, it is preferable to compute the predicted surface reflectance from atmospheric parameters that have already been band-integrated, rather than carrying out a stepwise band integration of the predicted surface reflectance as a function of atmospheric parameters. The use of an inappropriate method in the reverse computation can give rise to errors in surface reflectance prediction as large as 45 to 60 percent in certain spectral bandpasses.

The differences between predicted surface reflectances based on monochromatic as opposed to bandpass computations can be substantial in commonly used sensor bands, ranging from 2 to 22 percent in the examples considered here.

5. ACKNOWLEDGEMENTS

The author wishes to thank P.N. Slater and R.P. Santer for valuable discussions, T. Lloyd-Ball and G. Fedosejevs for technical assistance, and A. Kalil for typing the manuscript.

REFERENCES

- Palmer, J.M. and M.G. Tomasko, 1980, "Broadband Radiometry with Spectrally Selective Detectors", *Optics Letters*, 5:208-210.
- Royer, A., N.T. O'Neill, A. Davis, and L. Hubert, 1988, "Comparison of Radiative Transfer Models Used to Determine Atmospheric Optical Parameters From Space", *Proc. SPIE 928*, Orlando, Florida, in press.
- Tanré, D., C. Deroo, P. Duhaut, M. Herman, J.J. Morcrette, J. Perbos, and P.Y. Deschamps, 1986, "Simulation of the Satellite Signal in the Solar Spectrum", *Laboratoire d'Optique Atmosphérique, Université des Sciences et Techniques de Lille*, 59655 Villeneuve d'Ascq Cédex, France, 343 pages.
- Teillet, P.M., R.P. Santer, and P.N. Slater, 1989, "Surface Reflectance Retrieval Methods and the Effects of Spectral Shifts on Sensor Response", in preparation.
- Teillet, P.M., N.T. O'Neill, A. Kalinauskas, D. Sturgeon, and G. Fedosejevs, 1987, "A Dynamic Regression Algorithm for Incorporating Atmospheric Models into Image Correction Procedures", *Proc. IGARSS'87 Symposium*, Ann Arbor, Michigan, pp.913-918.

TABLE 1: Input conditions for the 5S code runs.

| | |
|-----------------------|----------------------------|
| Terrain elevation: | 0 km |
| Sensor altitude: | 999 km |
| Solar zenith angle: | 60° |
| Solar azimuth angle: | 150° |
| Sensor zenith angle: | 30° |
| Sensor azimuth angle: | 100° |
| Atmospheric profile: | Mid-latitude summer |
| Aerosol model: | Continental |
| Visibility: | 23 km |
| Surface reflectance: | Uniform vegetation |
| Spectral bands: | Landsat-5 TM (reflective) |
| | SPOT-1 HRV (multispectral) |
| | NOAA-9 AVHRR (reflective) |

TABLE 2: Results of 5S model runs. The forward case is based on the conditions specified in Table 1. The quotation marks on the input values are a reminder that the input is actually a spectrally varying $\rho(\lambda)$ and the tabulated ρ is merely the band-integrated value.

| Spectral Band | Forward 5S Run Input ρ | Reverse 5S Run A Output ρ_A | % difference between ρ and ρ_A | Reverse 5S Run B Output ρ_B | % difference between ρ and ρ_B |
|---------------|-----------------------------|----------------------------------|--|----------------------------------|--|
| TM1 | "0.1006" | 0.1015 | 0.89% | 0.1000 | 0.60% |
| TM2 | "0.1175" | 0.1181 | 0.51% | 0.1183 | 0.68% |
| TM3 | "0.0960" | 0.0954 | 0.63% | 0.0958 | 0.21% |
| TM4 | "0.5270" | 0.5270 | 0.00% | 0.5492 | 4.2% |
| TM5 | "0.3731" | 0.3743 | 0.32% | 0.6058 | 62.0% |
| TM7 | "0.2133" | 0.2142 | 0.42% | 0.2173 | 1.9% |
| HRV1 | "0.1170" | 0.1178 | 0.68% | 0.1174 | 0.34% |
| HRV2 | "0.1139" | 0.1144 | 0.44% | 0.1151 | 1.1% |
| HRV3 | "0.5275" | 0.5276 | 0.02% | 0.5449 | 3.3% |
| AVHRR1 | "0.1098" | 0.1107 | 0.82% | 0.1114 | 1.5% |
| AVHRR2 | "0.5156" | 0.5152 | 0.08% | 0.7521 | 46.0% |

TABLE 3: Differences between predicted surface reflectances based on bandpass versus monochromatic computations. The results are expressed in terms of percent difference between bandpass reflectance and monochromatic reflectance with respect to the bandpass reflectance values.

| Spectral Band | Percent Difference | Spectral Band | Percent Difference |
|---------------|--------------------|---------------|--------------------|
| TM1 | -1.9% | HRV1 | -1.8% |
| TM2 | -7.5% | HRV2 | -7.9% |
| TM3 | +3.3% | HRV3 | +1.6% |
| TM4 | +2.9% | AVHRR1 | +2.1% |
| TM5 | +8.8% | AVHRR2 | +21.8% |
| TM7 | +3.4% | | |

TABLE 4: Input conditions for 5S code runs used to study spectral bandpass radiance results using a limited number of monochromatic calculations.

| | |
|-----------------------|---|
| Terrain elevation: | 1.2 km |
| Sensor altitude: | 999 km |
| Solar zenith angle: | 60° |
| Solar azimuth angle: | 135° |
| Sensor zenith angle: | 0° |
| Sensor azimuth angle: | 101° |
| Atmospheric profile: | Mid-latitude summer |
| Aerosol model: | Continental |
| Visibility: | 100 km |
| Surface reflectance: | (i) uniform sand (ii) uniform vegetation |
| Spectral bands: | NOAA-9 AVHRR (reflective) |

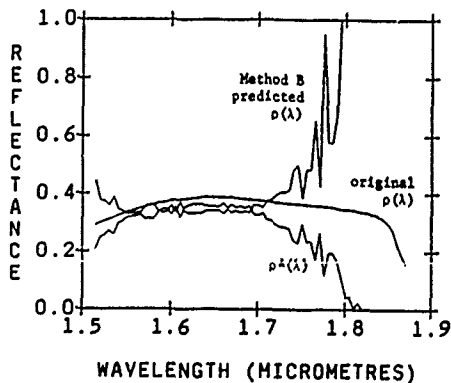


Figure 2: Vegetation reflectance inputs and outputs for the 5S code runs in the TM band 5 spectral region.

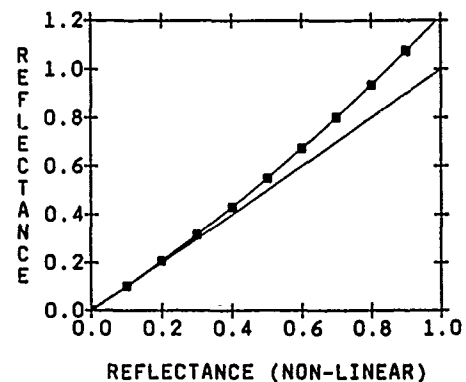


Figure 1: Comparison of predicted surface reflectance with (x axis) and without (y axis) taking multiple scattering into account. The points are based on reverse runs of the 5S code over vegetation for a wavelength of 0.480 micrometres and the curve is a quadratic fit. The straight line is the 1:1 line for visual reference.

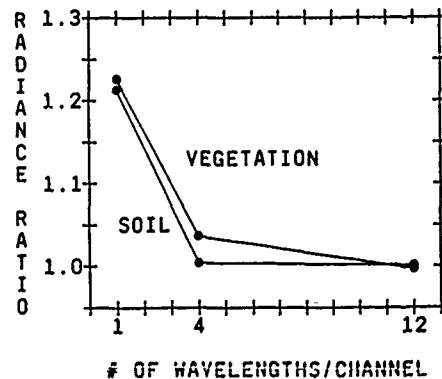


Figure 3: Comparison of apparent radiance values at sensor altitude in NOAA-9 AVHRR channel 2 generated by 5S runs at a limited number of wavelengths versus the full bandpass calculation. The radiance ratio is defined as the result from a limited number of wavelengths divided by the bandpass result.

Abstract

IMAGE BASED ATMOSPHERIC CORRECTION OF AIRBORNE IMAGERY

J. R. Freemantle†, J. R. Miller†, N. T. O'Neill‡, and A. Royer‡

†Center for Research in Experimental Space Science,
York University

4700 Keele Street, North York, Ontario, Canada, M3J 1P3

Tel: (416)-736-2100 x3510, Telex: 06524736, Fax: (416)-736-5516

‡CARTEL, Université de Sherbrooke

Sherbrooke, Québec, Canada J1K 2R1

Tel: (819)-821-7180, Fax: (819)-821-7238

The need for atmospheric correction of imagery for quantitative remote sensing has long been recognized. Unfortunately atmospheric measurements made coincident with image acquisition which would enable atmospheric correction are not always available. It therefore is desirable to be able to derive atmospheric parameters from the imagery alone in order to incorporate some sort of atmospheric correction procedure into the analysis of imagery.

For satellite image analysis image based darkest pixel algorithms have been used for a number of years over water. Over land the problem is more complex because the surface albedo is usually an unknown quantity. A number of methods of correcting airborne data over land have been tried which involve flying the aircraft at different altitudes and along different flight lines over the same target. These methods add expensive flight time to a remote sensing mission.

Using imagery with high spectral and high spatial resolution a number of new image based methods are being explored. They include using shadows in the imagery to calculate an irradiance ratio and oxygen absorption lines to determine aerosol optical depths. The results will be intercompared and evaluated with extensive ground truth measurements made concurrent with the aircraft overflight.

ANALYSE DES AEROSOLS À PARTIR DES IMAGES AVHRR-NOAA 9 ET 10. APPLICATION À LA CARTOGRAPHIE DES SULFATES.

Chartier, Lyne, O'Neill, Norman T. et Alain Royer

CARTEL, Centre d'applications et de recherches en télédétection.
Université de Sherbrooke, Sherbrooke, Québec, Canada. J1K 2R1

Résumé

Cette étude porte sur l'estimation des concentrations des aérosols à partir de luminances dans les bandes 1 et 2 des images AVHRR-NOAA au dessus des surfaces aquatiques et végétalisées. Un modèle d'inversion développé à partir du modèle 5S de Tanré *et al.* 1986, permet le calcul de l'épaisseur optique des aérosols (τ_a) sur tous les pixels de l'image. Une simulation de la variation des hypothèses du modèle en fonction de τ_a est réalisée. En exemple, une application du modèle d'inversion sur une image AVHRR-NOAA 10 est présentée. On y observe que les valeurs de τ_a se comparent aux données in-situ pour l'eau mais pas pour la végétation. Finalement, on montre qu'à partir de la répartition spatiale des aérosols de type continental, on peut estimer les concentrations en sulfates.

Abstract

In this paper, we present a method for estimating aerosol optical depth from the radiances of channels 1 and 2 of AVHRR-NOAA imagery over water and vegetation surfaces. An inversion model developed from the 5S model of Tanré *et al.* 1986, is employed to calculate the aerosol optical depth (τ_a) on a pixel by pixel basis. An analysis of the aerosol optical depth precision based on simulation studies of sensitivity to variations in water quality (chlorophyll concentration), water vapor content and vegetation surface reflectance was performed. The inversion model applied on an NOAA-10 AVHRR image shows that τ_a solutions are comparable to ground based estimates computed from extinction data over water surfaces only. Finally, assuming a continental type aerosol we compute the sulfate concentration from the aerosol optical depth maps.

Mots-clés: aérosols, modèle d'inversion, AVHRR-NOAA, sulfates

1.0 Introduction

L'imagerie AVHRR (Advanced Very High Resolution Radiometer) des satellites NOAA est déjà largement utilisée dans l'étude de l'évolution et de l'inventaire de la végétation. Toutefois, son utilisation se répand de plus en plus dans les études atmosphériques, en particulier pour le transport à longue distance des polluants (Chung, 1986). Le passage bi-journalier et l'importance du territoire couvert par ce satellite permet d'étudier de vastes régions et de visualiser simultanément certains phénomènes météorologiques significatifs à la dispersion des polluants.

Cette étude est orientée sur l'estimation de leur concentration, soit l'épaisseur optique des aérosols. Le but de cette recherche est la mise au point d'une méthode permettant d'estimer les concentrations des aérosols à partir des luminances dans les bandes 1 et 2 du visible et

du proche infrarouge. Etant donné que les sulfates, principaux polluants atmosphériques et parmi les plus importants optiquement, sont corrélés de façon significative au coefficient de diffusion de la lumière, on montre qu'il serait possible d'en déduire une cartographie quantitative des sulfates. Il faut toutefois noter que si le suivi des aérosols est possible au dessus des océans (ou de grands lacs dans notre cas) (Royer *et al.* 1988a), leur observation au dessus des continents est beaucoup plus difficile.

2.0 Méthodologie

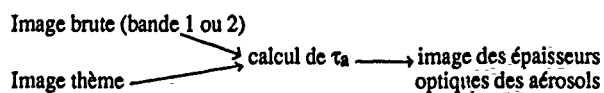
2.1 Modèle d'inversion

Le modèle d'inversion utilisée dans l'étude est dérivé du modèle de transfert radiatif 5S (Tanré *et al.* 1986). Une modification importante a été apportée, soit le calcul de l'épaisseur optique des aérosols en fonction de l'humidité relative dans l'air. L'humidité relative ayant un effet sur la taille et l'indice de réfraction de ces particules, les paramètres optiques reliés aux aérosols s'en trouvent modifiés.

Le calcul des épaisseurs optiques des aérosols (τ_a) s'effectue par itération sur l'égalité (à 10^{-5} près) entre la réflectance apparente mesurée et celle calculée par le modèle en supposant connue la réflectance au sol (végétation ou eau). Tous les τ_a solutions forment l'image des épaisseurs optiques des aérosols. La précision du modèle a été discutée par ailleurs (Royer *et al.* 1988a). En particulier, nous avons examiné la précision de l'inversion en fonction des angles de visée qui peuvent être importants dans le cas d'une image AVHRR-NOAA ($\pm 50^\circ$).

2.2 Traitement des images

La plupart des études effectuées sur la mesure des aérosols ne sont réalisées que sur un type de surface terrestre homogène à la fois, par exemple la forêt (Tsonis, 1987) ou l'eau (Freemantle, 1985). Le modèle développé ici considère les surfaces végétalisées et aquatiques:



Ce schéma présente le principe du traitement des images. L'image brute est une image AVHRR-NOAA 9 ou 10 et la bande est choisie en fonction de la cible considérée. S'il s'agit de végétation nous utilisons la bande 1 englobant l'absorption de la chlorophylle et pour laquelle la réflectance est relativement stable; si c'est de l'eau nous utilisons la bande 2 correspondant à une réflectance nulle. Nous combinons en sortie les deux bandes pour n'avoir qu'une image de τ_a , ramenée à une longueur d'onde de $0.55 \mu\text{m}$. L'image thème obtenue par un calcul de l'indice de biomasse $(\text{IR}-\text{R})/(\text{IR}+\text{R})$ permet

de séparer les surfaces végétalisées des surfaces aquatiques. Un découpage de l'histogramme de cette image est réalisé pour identifier les surfaces végétalisées, aquatiques et non-classifiées.

3.0 Résultats

3.1 Analyse de la réflectance apparente en fonction de la réflectance au sol

Les figures 1a et 1b montrent la variation de la réflectance apparente intégrée sur toute la bande AVHRR en fonction de la réflectance au sol pour 4 valeurs de τ_a . Il existe un point critique où la réflectance apparente sera toujours la même quelque soit la valeur de τ_a (Kaufman, 1987). On observe donc une zone défavorable où l'inversion est très instable, $0.1 < \rho < 0.25$ pour la bande 1 et $0.08 < \rho < 0.2$ pour la bande 2. Ces réflectances correspondent typiquement au sol nu. Par contre, la végétation dans la partie rouge du spectre ($0.6-0.7 \mu\text{m}$) pourrait constituer des zones de référence au dessus de la terre. En effet, les surfaces végétales qui couvrent près de 50% des continents ont une réflectance intégrée sur la bande 1 AVHRR assez faible, de l'ordre de 3-4%, et restent relativement stables ($\pm 2\%$) (Royer et al., 1988b; Kaufman et Saundra, 1988).

3.2 Variabilité du signal aux hypothèses du modèle

Nous avons fait varier différents paramètres d'entrée du modèle d'inversion pour évaluer la précision des τ_a obtenus suite au traitement des images. A la figure 2a, on a montré l'effet de la variation de la concentration en chlorophylle dans l'eau pour la bande 1 si l'on prend la réflectance de l'eau comme référence (modèle dérivé de Deschamps et al., 1980). Quand la concentration en chlorophylle varie de $.05 \text{ mg/m}^3$ nous obtenons une précision de 10% sur τ_a . Pour la végétation, qui est plus variable en réflectance que l'eau, nous obtenons une précision moins bonne (figure 2b). Quand la réflectance varie de $.02$ nous augmentons l'erreur sur τ_a par un facteur 2. Toutefois ce facteur tend à diminuer quand τ_a augmente. Dans la bande 2 (figure 2c), la variation de la quantité de vapeur d'eau modifie la réflectance apparente mesurée (absorption). L'erreur sur les valeurs de τ_a est de moins de 10% quand la quantité de vapeur d'eau varie de $.3 \text{ g/cm}^2$.

3.3 Exemple d'application du modèle d'inversion sur une image AVHRR-NOAA

Le modèle a été appliqué à une image AVHRR-NOAA 10 du lac Ontario du 11 août 1987. La figure 3 montre le résultat de la combinaison des images des épaisseurs optiques des aérosols de la bande 1 (végétation) et la bande 2 (eau). On remarque que τ_a au dessus de l'eau est plus faible qu'au dessus de la végétation. Nous observons également une bordure sur le pourtour du lac correspondant aux pixels mixtes dus à la contamination du signal de l'eau par l'effet d'environnement et la réflexion du fond. Les pixels non-classifiés représentent des zones de sol nu, urbaines ou des nuages. En comparant les données in-situ aux données image, on voit que pour les secteurs près du lac ou sur le lac, les valeurs sont du même ordre de grandeur à 10% près. Au dessus des surfaces végétalisées, le modèle d'inversion surestime beaucoup l'épaisseur optique des aérosols.

4.0 Discussion

D'après les différents tests effectués, nous considérons que le modèle d'inversion développé ici est assez rapide et précis et que son utilisation pour déterminer l'épaisseur optique des aérosols est avantageuse. Toutefois, l'inversion au dessus des surfaces végétalisées demande à être étalonnée puisque l'écart est trop important par rapport aux mesures au sol et par rapport aux zones

voisines situées au dessus de l'eau avec la valeur de ρ végétation utilisée (0.04). La figure 2b montre qu'il existe une valeur de ρ (0.06) pour laquelle on a des valeurs τ_a homogènes sur toute l'image (eau/végétation).

Application à la cartographie des sulfates

La détermination de la répartition spatiale des épaisseurs optiques des aérosols nous permet d'envisager la cartographie des concentrations en sulfates. En effet, il existe une relation entre le coefficient de diffusion des aérosols et la concentrations des sulfates au sol (tableau 1). Etant donné que τ_a représente l'intégration verticale du coefficient de diffusion, on peut dire que ce paramètre est corrélé à la concentration intégrée des sulfates. Si l'épaisseur de la couche des aérosols est sensiblement constante dans une variété de conditions on peut calculer la concentration de sulfates au sol. La concentration des sulfates estimée, pour la région de Toronto, à partir des données images au dessus de l'eau est de $4.8-7.8 \mu\text{g/m}^3$ en supposant une épaisseur de couche des aérosols (scale height) de 1 km et un modèle des aérosols de type continental; et celle mesurée in-situ par l'OME pour une journée est de $7.5 \mu\text{g/m}^3 (\pm 1.7)$.

5.0 Conclusion

La méthode proposée pourrait donc permettre de cartographier les sulfates sur les zones continentales urbanisées. Le principal handicap rencontré concerne la modélisation de la végétation par rapport aux bandes très larges du AVHRR-NOAA. En effet, la bande passante du visible (bande 1) chevauche un peu le décalage de "l'épaule rouge" des réflectances dans le proche IR, la valeur intégrée sur cette bande peut donc varier de ± 0.02 suivant le type de végétation (agriculture vs forêt par exemple). L'arrivée de Landsat 6 avec le capteur SEAWIFS ayant une résolution équivalente à NOAA mais avec des bandes spectrales plus étroites pourra améliorer les calculs de l'épaisseur optique des aérosols au dessus des surfaces végétalisées.

Ce projet a été en partie subventionné par le Ontario Ministry of the Environment (OME Project #349G) et par le CRSNG, Canada (A8643 et A1765)

6.0 Bibliographie

1. Chung, Y.S. "Air pollution detection by satellites: the transport and deposition of pollutants over oceans", Atmospheric Environment, Vol. 20, p.617-630, 1986.
2. Deschamps, P.Y., Tanré, D. et Viollier, M. "Evaluation critique des exigences radiométriques pour un équipement mesurant à bord d'un satellite la couleur des océans", Contrat ESA 3986/79/F/CG, 1980.
3. Freemantle, J.R., Miller, J.R. "Atmospheric Aerosol Variability over the Great Lakes from NOAA-7 AVHRR imagery", 9th Canadian Symposium on Remote Sensing, St-John, NFLD, p.263-270, 1985.
4. Kaufman, Y.J., Saundra, C. "Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery" Int. Jour. of Remote sensing, Vol. 9, p.1357-1387, 1988.
5. O'Neill, N.T., Royer, A., Hubert, L., Freemantle, J. "Sensibilité des luminances satellitaires aux variations des aérosols et en particulier des sulfates", IGARSS89/12ème Symposium canadien sur la télédétection, Vancouver, Canada, 1989a.
6. O'Neill, N.T., Royer, A., Hubert, L., Miller, J.R., Freemantle, J., Austin, G., Davis, A. "Critical Evaluation of Atmospheric Pollutant Parameterization from Satellite Imagery", Report presented to the Ontario Ministry of the Environment, Project No. 349G, 1989b.
7. Royer, A., Chartier, L., Côté, P. "Analyse de la contribution de l'atmosphère à la réflectance de l'eau dans les deux premiers canaux AVHRR des satellites NOAA", 4ème Colloque sur les signatures spectrales d'objets en télédétection, Aussois, 18-22 janvier 1988, 6 p., 1988a.

8. Royer, A., O'Neill, N.T., Hubert, L. "Comparison of Radiative Transfer models used to determine Atmospheric Optical Parameters from space", SPIE'S 1988, Technical Symposium on Optics and Sensors, Orlando, Florida, U.S.A., April 4-8 1988, Proc. Vol. 928, 18 p., 1988b.
9. Tanré, D., Deroc, C., Duhaut, P., Heaman, M., Morcette, J.J., Perbos, J., Deschamps, P.Y. "Simulation of Satellite Signal in the Solar Spectrum (SS). Logiciel et Documentation", Laboratoire d'Optique Atmosphérique, (LOA), Lille, France, 148 p., 1986.
10. Tsionis, A.A. "A method for Inferring Aerosol Properties from Satellites Data over Forested Terrain", Atmospheric Environment Vol. 21, No.11, p.2467-2471, 1987.

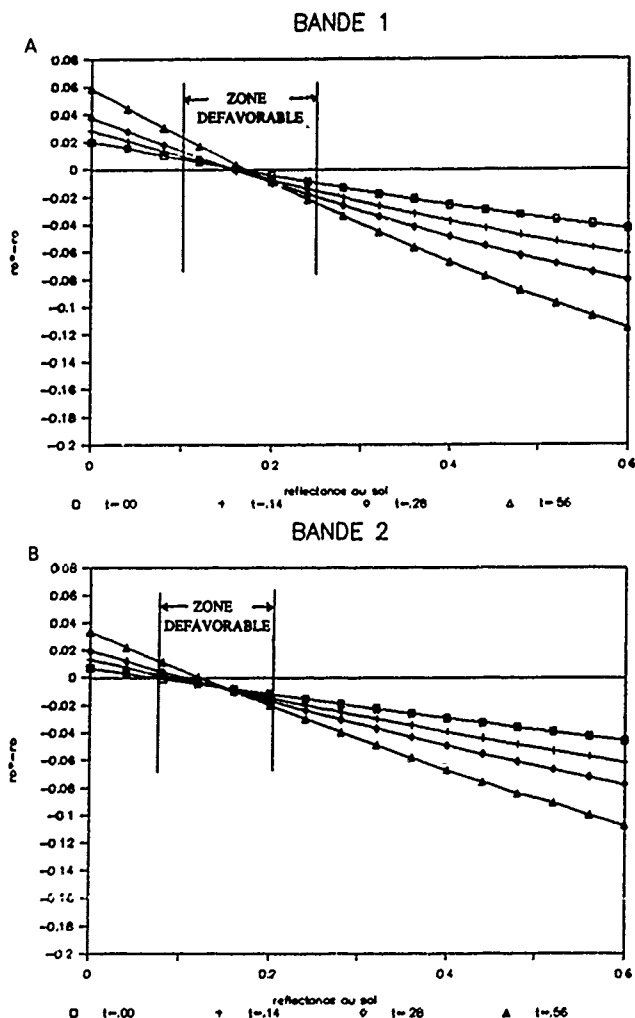


Figure 1. Variation de la réflectance apparente (ρ^a) en fonction de la réflectance au sol pour 4 valeurs de τ_s . Simulation AVHRR-NOAA 10, angle zénithal solaire 45° , visée au nadir, aérosols de type continental standard; ρ^a : réflectance apparente, ρ_s : réflectance au sol.

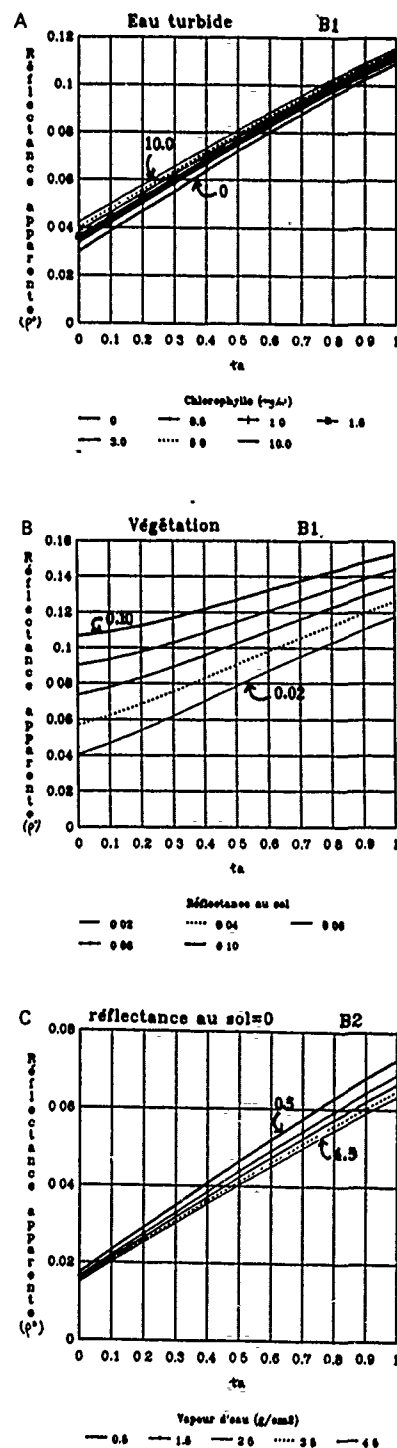


Figure 2. Variation de la réflectance apparente en fonction de τ_s pour différents paramètres d'entrée : (a) chlorophylle, (b) végétation, (c) vapeur d'eau. Simulation AVHRR-NOAA 10, visée au nadir, aérosols de type continental, atmosphère de latitude moyenne d'été.

REMOVAL OF THE SCAN ANGLE EFFECT ON NOAA-AVHRR DATA IN VISIBLE AND NEAR-IR PARTS OF THE ELECTROMAGNETIC SPECTRUM

Chongjun YANG - Alain VIDAL - Laboratoire Commun de Télédétection CEMAGREF-ENGREF - B.P. 5095 - 34033 MONTPELLIER
CEDEX 1 FRANCE

Damien LEPOUTRE - GEOSYS 3 Rue HERMES Parc Technologique - 31520 TOULOUSE - FRANCE

Denis POZZOBON - ORMVAG B.P. 79 - KENITRA - MOROCCO

ABSTRACT

NOAA-AVHRR data are associated to a great extent with the scan angle effect. It is necessary to remove the scan angle effect from this type of data prior to analysis. This paper provides a simplified and feasible approach to correct scan angle effect for AVHRR-NOAA data in the visible and near-IR parts of the electromagnetic spectrum (channels 1 and 2). Corrections attempt to provide pixel by pixel nadir responses. The approach is based on a simplified scene spectral radiance equation that is developed in this study. The values of relevant atmospheric parameters such as the total normal optical thickness, and path spectral radiance for each spectral band were estimated from the real images. Lambert's law was assumed as a bidirectional reflection at the ground surface. The contributions due to multiple reflection were neglected. The feasibility and the limits of this approach have been tested and reviewed against the imagery from AVHRR-NOAA 9 above a pasture of Aveyron in the south of France.

Key-words : NOAA-AVHRR Data, Scan angle effect, view zenithal angle, Radiometric Correction.

1 - INTRODUCTION

Since 1979 the AVHRR (Advanced Very High Resolution Radiometer) sensors have been installed on the NOAA satellites, there have been a number of studies in the use of NOAA-AVHRR Data such as to monitor the herbaceous biomass production, provide crop condition assessments, detect tropical deforestation, make vegetation classification, monitor seasonal and phenological changes in vegetation on a global scale, and monitor drought impact on forage surface and many other examples.

The NOAA-n satellites have a sunsynchronous polar orbit of about height 830 km (± 18 km), with a resolution of 1.1 km at nadir. The NOAA-AVHRR sensor scans 1024 pixels on either side of nadir (approximately 1350 km to the east and to the west of nadir). Scan angles increase from zero to 55.4 degrees on both sides of nadir. This situation is meant by a possible variation of a view zenithal angle range of $\pm 67^\circ$. Thus, the spectral radiance falling into the NOAA-AVHRR sensor must pass through more of the earth's atmosphere as the scan angle increases (JOHNSON et al. 1986). In fact, about 255 pixels to the left and right of nadir (the center

25 % of the scanline) can be better used without scan angle correction. That is, the 75% portion of the NOAA-AVHRR data is strongly contaminated by the scan angle effect.

On the other hand, a comparative analysis of NOAA-AVHRR data taken on a series of dates can provide additional elements of the identifying signature because many types of features exhibit changes with the passage of time. It also follows that more information about the earth resource of an area can be obtained from imagery of it that have been taken on several or more dates than from imagery of it taken on only one date. For example, the normalised difference vegetation index between the near infrared and the visible channels of NOAA-AVHRR is a important tool to follow up the evaluation of drought impact on fodder production. However, such a use always needs a series of satellite data collected during some different dates. It means that the multitemporal data measured by the NOAA-AVHRR sensor at the different scan angles, for the same ground location are used. It is impossible to extract information for the same target at the earth's surface from the NOAA-AVHRR data without scan angle correction.

An approach is proposed in providing answers to the above described necessity. Scan angle correction was made for AVHRR-NOAA data in the visible and near-IR parts of the electromagnetic spectrum (channels 1 and 2). Each pixel value along the scanline was corrected to simulate a nadir-response. The approach is based on a simplified scene spectral radiance equation that is developed in this study. The values of relevant atmospheric parameters such as the total normal optical thickness, and path spectral radiance for each spectral band were estimated from the real images. Lambert's law was assumed as a bidirectional reflection at the ground surface. The contributions due to multiple reflection were neglected.

2 - MODELLING

In this section, we have an interest in the derivation of the correction model that transforms the spectral radiance measured by the NOAA-AVHRR sensor at the arbitrary view zenithal angle (corresponding to the scan angle) to the spectral radiance at the zero view zenithal angle, that is only a function of the solar radiation, the atmospheric characteristics, the angles of incidence, the wavelength and surface material. In

all the following expressions, the wavelength subscriptions have been omitted.

The relation between the view zenithal angle and the scan angle is plotted in figure 1. Θ and α are the view zenithal angle and the scan angle respectively (rad). R (km) is the radius of the Earth, H (km) the orbital height. We have a following transformation form of the scan angle to the view zenithal angle

$$\sin \Theta = \frac{R+H}{R} \sin \alpha \quad (1)$$

Thus, in our model, the view zenithal angle correction will be practically discussed, instead of the scan angle correction that was used in the literature.

A simple sketch is depicted in figure 2. The spectral radiance reaching the NOAA-AVHRR sensor consists of two parts: solar electromagnetic radiation scattered by the atmosphere in the direction of the sensor including the reflected radiation part from the environment surface of the target, and solar electromagnetic radiation reaching the target of the earth (the sum of the irradiances of the direct sunlight and of the skylight) and reflected by the target in the direction of the sensor. The target surface is assumed as a horizontal flat lambertian surface on a macroscopic scale. The atmosphere is treated as a series of horizontally homogeneous parallel slabs. The skylight was treated as a uniform hemispherical source. The direct sunlight was treated as a well-collimated source (parallel rays). The variation of the path spectral radiance with view zenithal angle is neglected. Let L_a ($W.m^{-2}.sr^{-1}.\mu m^{-1}$) be the total atmosphere upward scattered path spectral radiance including the radiance reflected by the ground (the environments of the target). The total spectral radiance L_{Θ} ($W.m^{-2}.sr^{-1}.\mu m^{-1}$) recorded by the sensor may be represented as

$$L_{\Theta} = \frac{p}{\pi} (e^{-\tau \sec \Theta} E_0 \cos S + E_d) e^{-\tau \sec \Theta} + L_a \quad (2)$$

where S (rad) is the solar zenithal angle, Θ the view zenithal angle, E ($W.m^{-2}.\mu m^{-1}$) the solar irradiance at the "top" of the atmosphere (the extraterrestrial solar irradiance), E_d ($W.m^{-2}.\mu m^{-1}$) the diffuse sky irradiance, p (dimensionless) the target reflectance and τ the total normal optical thickness, which equals the sum of the separate optical thicknesses of all the attenuating constituents. All variables are specific to each wavelength band except the solar and view zenithal angle terms, S and Θ . Equation (2) may be rewritten in the form

$$L_{\Theta} = \quad (3)$$

$$\frac{p}{\pi} (e^{-\tau \sec \Theta} E_0 \cos S + E_d) e^{-\tau \sec \Theta} + \tau + L_a$$

When the view zenithal angle Θ is zero, the value of the parameter $\sec \Theta$ becomes one. The above equation becomes

$$L_{0^\circ} - L_a = \frac{p}{\pi} (e^{-\tau \sec \Theta} E_0 \cos S + E_d) e^{-\tau} \quad (4)$$

Combining equations (3) and (4), we have

$$L_{\Theta} = (L_{0^\circ} - L_a) e^{-\tau \sec \Theta} + \tau + L_a$$

that is

$$L_{0^\circ} = (L_{\Theta} - L_a) e^{\tau \sec \Theta} + L_a \quad (5)$$

The above equation presents the relation between the spectral radiance measured by the satellite sensor at the arbitrary view zenithal angle and the simulated spectral radiance at the zero view zenithal angle (YANG 1986).

3 - APPLICATION

3.1. Algorithm of transformation from radiance to digital numbers

The NOAA-AVHRR sensor is not monochromatic, but measures integrated upwelling space radiance within some broad wavelength bands. The relation between the mean spectral radiance L recorded by the sensor and digital numbers DN (dimensionless) in the given band is quantified by the following expression

$$DN = A \cdot L + B \quad (6)$$

where A , B are the calibration coefficients. Let DN_{0° , DN_{Θ} , DN_a be the digital numbers of the corresponding mean L_{0° , L_{Θ} , L_a , respectively. Equation (5) can be rewritten as

$$DN_{0^\circ}/A - B/A =$$

$$(DN_{\Theta}/A - B/A - (DN_a/A - B/A)) e^{\tau \sec \Theta} + DN_a/A - B/A$$

that is

$$DN_{0^\circ} = (DN_{\Theta} - DN_a) e^{\tau \sec \Theta} + DN_a \quad (7)$$

If all the values of the parameters on the right of equation (7) are known, DN_{0° can be calculated. The DN_{Θ} is directly available from NOAA-AVHRR data.

3.2. Estimation of the total normal optical thickness

The value of the total normal optical thickness can be usually evaluated from Lowtran 5 (or Lowtran 6) Code with the assumptions of a mid-latitude summer model and a rural aerosol model with a horizontal visibility of 23 km (corresponding to a clear sky). An algorithm is proposed for estimating the total normal optical thickness from real imagery. As shown in figure 3, let the targets 1 and 2 be the same nature (the surface material needs to be identified). According to equation (7), we have

$$DN_{0^\circ 1} = (DN_{\Theta 1} - DN) e^{\tau \sec \Theta 1} + DN \quad (8)$$

and

$$DN_{0^\circ 2} = (DN_{\Theta 2} - DN) e^{\tau \sec \Theta 2} + DN \quad (9)$$

On the assumptions that have brought about equation (5), (8), and (9), the calculated value $DN_{0^\circ 1}$ should be equal to the $DN_{0^\circ 2}$, that is

$$(DN_{\Theta 1} - DN) e^{\tau \sec \Theta 1} + DN =$$

$$(DN_{\Theta 2} - DN) e^{\tau \sec \Theta 2} + DN \quad (10)$$

Being derived from equation (10), the total normal optical thickness becomes

$$\tau = \frac{\ln((DN \ominus_1 - DN) / (DN \ominus_2 - DN))}{\sec \ominus_2 - \sec \ominus_1} \quad (11)$$

If K numbers of couple points selected like the points 1 and 2 are found, the total normal optical thickness can be given by

$$\tau = \frac{1}{K} \sum_{i=1}^K \tau_i \quad (12)$$

3.3. Estimation of the path radiance

Four methods of estimating the combined path radiance and sensor offset terms in sensor-measured radiance values have been described previously in the literature (ROBERT E. CRIPPEN 1987). They are (1) atmospheric measurements and modelling combined with knowledge about the sensor calibration, (2) dark-pixel subtraction, (3) radiance-to-reflectance conversion and (4) the regression method. Dark-pixel subtraction, also known as the histogram minimum method, is a pixel with zero illumination or additive components. In this study, DN^a was estimated by using the minimum digital number in the corresponding band.

3.4. Computation of the view zenithal angle

A view zenithal angle \ominus can be computed by using the pixel location (pixel number) of given target along the scanline. The corresponding equation is

$$\ominus =$$

$$\arcsin((1+H/R) \cdot \sin((N_p - n/2) \cdot 110.8 \cdot \pi / (180 \cdot n))) \quad (13)$$

Where the n is the total pixel number per line, N_p the local pixel number, H the orbital height, and R the radius of the earth.

Up till now, the algorithm of all the parameters needed in equation (7) has been proposed, which makes our approach of scan angle correction more applicable.

3.5. Data and results

Applications have been realized for two imagers of NOAA-9-AVHRR, in visible and near infrared channels, respectively taken on 11/10/1985 and 13/10/1985. The study area is located in the AVEYRON in the south of France. The area (360 km² 420 km) is dominated by the fodder production. The latitude and the longitude of the area centre is 44°21'N, 2°36'E, respectively. The view zenithal angle of the area centre is 33°20' for the image taken on 11/10/85, 53°48' for the image taken on 13/10/85. The NOAA-AVHRR data used in this study are received by the Centre de Météorologie Spatiale, Lannion, France. The data were resampled by a closest neighbour procedure and rectified to a Lambert projection map using the satellite ephemerides (D. POZZOBON 1987). After geometric correction, location accuracy is within 1 km. The minimum digital numbers, chosen as the values of DN^a in equation (7), of the image taken on 11/10/85 were 18, 11, corresponding respectively to the channel 1 and channel 2. 24, 15 correspond to the image taken on 13/10/85. The results are shown as figure 4 and figure 5.

4 - CONCLUSION

Based on this study, the conclusions are listed below :

- 1° the radiance measured by the NOAA-AVHRR sensor includes a very strong scan angle effect.
- 2° it seems feasible to remove the scan angle effect from the NOAA-AVHRR data in visible and near-IR channels by using our model.
- 3° NOAA-AVHRR data must first be subjected to certain geometric correction prior to the above correction.
- 4° The normalized difference vegetation index is often utilized to minimize the influence of atmospheric conditions that may vary from day to day or over longer intervals. But, the scan angle effect can not be removed by this usual technique.
- 5° The test of the proposed method will be taken into account later on.

REFERENCES

- ROBERT, E., C., 1987 The regression intersection method of adjusting image data for band ratioing, INT.J.Remote Sensing. VOL.8(2), 137-155.
- JOHNSON, W.R., BOATWRIGHT, G.O., and LIN, C.C., 1986 Models to correct scan angle effects for the descending NOAA-6 and NOAA-8 satellite sensors. Lockheed engineering and management services company 7500 greenway center drive Suite 900 Greenbelt, Maryland 20770.
- YANG, C.J., 1986 Prise en compte de l'angle d'observation dans correction radiométrique sur l'image NOAA. Rapport de DEA, UPS-CERAGREF, TOULOUSE, FRANCE.

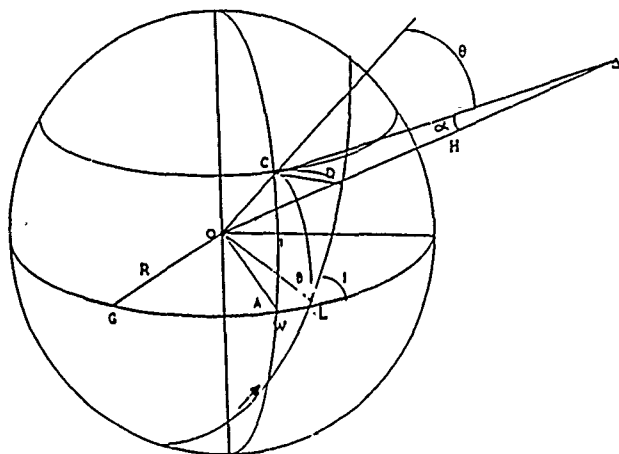


FIGURE 1. Geometric relation between the view zenithal angle θ and the scan angle α .

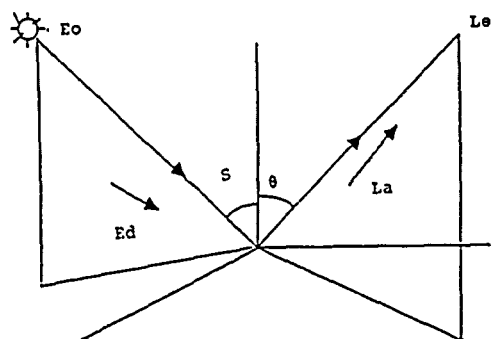


FIGURE 2. Reflected solar radiance and path radiance measured by a scanner.

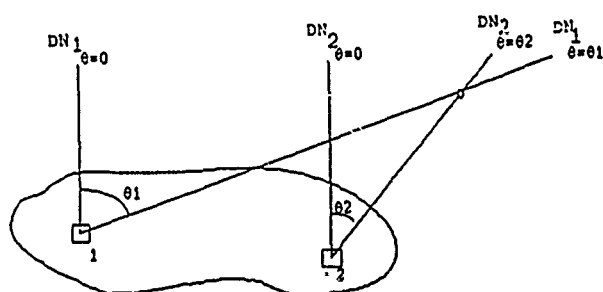


FIGURE 3. The target 1 and 2 are identified in the reflectance factor. The total normal optical thickness is estimated from the values of θ_1 , θ_2 , DN , DN_{θ_1} , DN_{θ_2} .

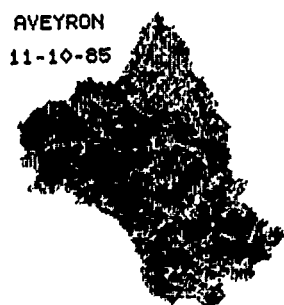


FIGURE 4a. Normalised difference vegetation index between the channel 1 and channel 2 of NOAA-AVHRR data taken on 11/10/85.



FIGURE 4b. Normalised difference vegetation index between the channel 1 and channel 2 of NOAA-AVHRR data taken on 11/10/85, corrected using equation (7).

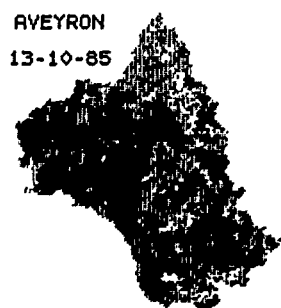


FIGURE 5a. Normalised difference vegetation index between the channel 1 and channel 2 of NOAA-AVHRR data taken on 13/10/85.



FIGURE 5b. Normalised difference vegetation index between the channel 1 and channel 2 of NOAA-AVHRR data taken on 13/10/85, corrected using equation (7).

A MEASUREMENT PROGRAM FOR THE VALIDATION OF ATMOSPHERIC RADIATIVE TRANSFER CODES

R.P. GAUTHIER, F.J. AHERN, P.M. TEILLET

CANADA CENTRE FOR REMOTE SENSING
2464 SHEFFIELD ROAD, OTTAWA, ONTARIO.

ABSTRACT

A field campaign was undertaken in the summers of 1987 and 1988 in order to provide measurements of atmospheric parameters which could subsequently be used to validate a radiative transfer model for the purposes of atmospheric correction of remotely sensed images. The main objective was the validation of a radiative transfer model through the measurement of the model inputs and outputs. The details of the field measurements are discussed pointing out problem areas in the determination of optical depths at the 1% level. The measurements in the range from 390nm to 1100nm were made with a spectroradiometer having a spectral resolution of 6nm. The full spectral nature of the optical depth data reveals the difficulties of assessing a true "continuum" level for the determination of the aerosol scattering due to the overlapping of the gaseous absorption. Also, the behaviour of the gaseous absorption can be studied and correlated with sky conditions at the time of the observations. The determination of the Langley plot intercept spectrum has been made to an accuracy of about 0.5%, reflecting the high quality of the full spectral data.

1. INTRODUCTION

The demand for higher accuracy in reflectance retrievals from remotely sensed data in the contexts of forestry, bathymetry, global change studies and sensor calibration, among others, was the motivation for undertaking a large program whose main objective was to provide an atmospheric correction capability for image production and analysis systems (both research and production oriented) to an accuracy of 5% or better. While this order of accuracy will be difficult to achieve in practice, it is realistic to consider it as a target in order that the difficulties presented by such a figure can be studied in detail. Correction of remotely sensed data for the effects of the intervening atmosphere has two aspects. The first is the radiance contributed to the scene by the scattering of the downwelling irradiance by the atmospheric constituents between the sensor aperture and the surface. The second is the attenuation of the upwelling surface radiance by the atmospheric constituents between the surface and the sensor aperture. A standard technique for removing the first component, the path radiance, is the use of reference targets in the scene. For a target of known reflectance, the path radiance contained in the sensor signal can be easily obtained and subtracted from the rest of the data. The second component, the atmospheric

transmittance, must be calculated with a radiative transfer model, since it is a characteristic of the interaction of the upwelling scene radiance and the intervening atmosphere which can not be directly measured. The path radiance as obtained from a reference target in an image can serve as a measure of the specific atmospheric conditions pertaining to a particular scene for which a radiative transfer calculation can be made in order to obtain the transmittance. The present work focusses on the validation of a radiative transfer model for performing this calculation for data obtained from both satellite and airborne sensors.

2. THE MEASUREMENT PROGRAM

The validation of radiative transfer models involves measuring both the inputs and outputs (i.e. the properties of the atmosphere and ground reference target) for a variety of atmospheric conditions both outside of and during the overpass of a satellite and/or airborne sensor. Using measured inputs, various radiative transfer models can be intercompared as to how well they predict the measured outputs. The model which most consistently predicts results closest to the observations is then chosen and can be implemented in image production and analysis systems in the most efficient manner. The measured inputs to the models included the optical depth of the various atmospheric components (gaseous, aerosol, molecular) as well as the reflectance function of the reference ground target. The measured outputs included the global and sky downwelling irradiance, the sky radiance at various different locations and the path radiance over the reference target. This latter quantity was measured from both an airborne platform and satellite platforms. The measurement program was performed near Ottawa at the Gatineau Satellite Station and on neighbouring Grand Lake in 1987 and was relocated to the Baskatong Reservoir (some 250km north of Ottawa) in 1988 since the reservoir offered a much larger surface area for the airborne work (i.e. free of surrounding terrain effects). The measurement program included observations to permit:

- atmospheric corrections for off-nadir viewing sensors (e.g. SPOT HRV, MEIS II);
- extension of wavelength range for atmospheric corrections into the short wave infrared region (out to 2.5 μ);
- inclusion of polarisation effects in atmospheric corrections.

2.1 Radiometry of the Downwelling Field

The total atmospheric optical depth was measured using the Langley plot technique. The various sources of systematic error inherent in this method are discussed by Reagan et al. (1986). The instrumentation used consisted of a LI-COR spectroradiometer on an altazimuth mount operating over the spectral range from 390nm to 1100nm with a resolution of 6nm. A filter wheel radiometer operating out to 3.5 μ was mounted with the LI-COR to permit discreet bandpass (20nm) observations in TM short wave infrared bandpasses. The sensor apertures were bore-sighted with a solar viewing tube to ensure that the solar disc was centered on the sensor apertures during the observation. The LI-COR was used to make the downwelling irradiance (total and sky) and sky radiance measurements. These latter quantities were not measured with the filter wheel radiometer, due to an unacceptably low signal-to-noise for the observations. The global downwelling irradiance was measured by pointing the LI-COR to the zenith and taking a scan of the light incident on a diffuser. The sky irradiance was measured by shading the diffuser and taking a scan. The sky radiance was measured over a grid formed by almucanters evenly spaced in elevation from 30 degrees above the horizon to the zenith. These 'sky maps' were made with and without polarisers in order to measure the state of polarisation of the incident radiation field.

The instrument calibration was established by extrapolating the Langley plots for 7 clear half days in 1987 and 13 clear half days in 1988 to zero air mass and then taking the average of these independent estimates of the exo-atmospheric response. The standard deviation of the mean was less than 1% over the range from 390nm to 1100nm, and less than 0.5% over most of that range for both observing seasons. Total optical depths for each individual observation were then calculated from

$$\delta = \frac{(\ln(Q) - \ln(\langle Q_0 \rangle))}{A}$$

where

- δ is the total optical depth,
- Q is the quantized signal level at each wavelength for each observation,
- $\langle Q_0 \rangle$ is the mean exo-atmospheric signal level,
- A is the air mass for the observation.

An example of an optical depth spectrum can be seen in Figure 1. The most striking features are the rapidly increasing attenuation towards the blue due to Rayleigh scattering and the strong attenuation in the various absorption features, mostly due to O_2 and H_2O . The relatively shallow, broad O_2 absorption of the Chappuis band is not apparent in the total optical depth spectrum, but becomes obvious when the Rayleigh component is subtracted as shown in Figure 2.

The Rayleigh optical depth is supposed to be a "well-known" quantity if local pressure and temperature are taken into consideration. However, an examination of the literature indicates absolute differences of up to 4% in published computations, tabulations and parameterisations (Teillet, private communication), primarily because of differences in the depolarisation factor and atmospheric index of refraction used by various investigators. Thus, the Rayleigh correction must be carefully implemented if optical depth determinations are to be made at the 0.5% level. The Rayleigh subtracted optical depth spectrum reflects the attenuation characteristics of the gaseous absorption and aerosol absorption and scattering. The aerosol component is the most variable

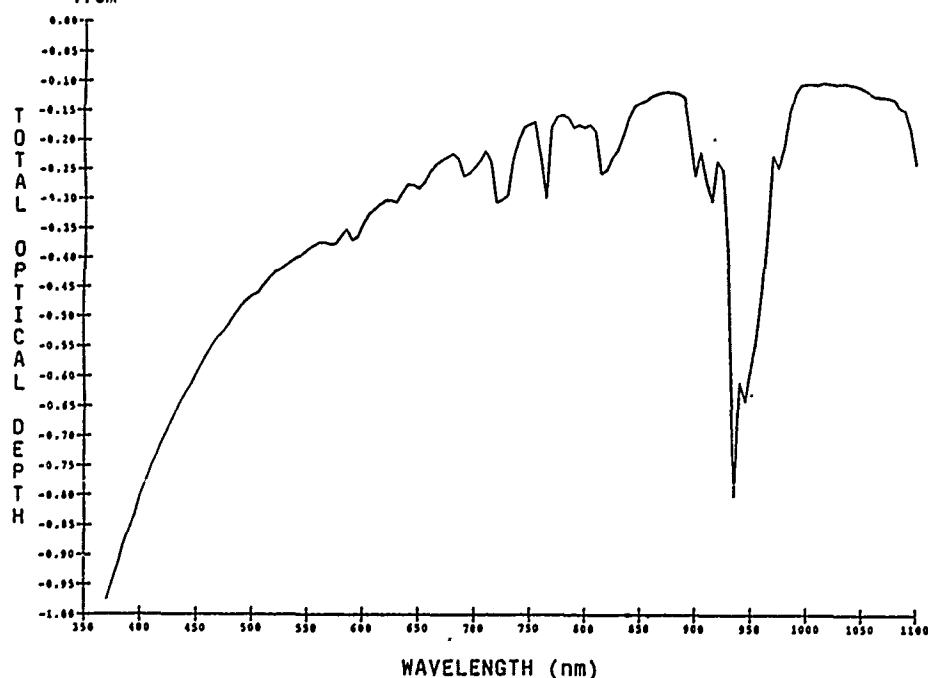


Figure 1:

Total optical depth spectrum exhibiting typical Rayleigh behaviour towards the blue.

(spatially and temporally) and is subsequently the most crucial component in the determination of the composition of the atmospheric constituents for the purposes of doing the radiative transfer calculation. The primary interaction between the aerosols and the irradiance is scattering and so the attenuation is of a broad spectral nature. It is usually described as a power-law $\delta = \alpha\lambda^\beta$. In order to determine the coefficients α and β , it is necessary to determine points in the spectrum which are produced by this broadband aerosol scattering, so-called "continuum" points. The primary difficulty in doing so is the presence of the gaseous absorption features which overlay the true continuum and produce a false or "pseudo-continuum". It is important to be able to identify all gaseous absorption features so that true continuum points may be found. The full spectral data reveals that such points are far fewer than has been assumed from filter wheel data since the gaseous absorption is so dominant. The regions least affected by absorption must be located and used consistently. The fact that not all the absorbing species are identified also makes this a difficult task, and examination of many spectra is necessary to determine which areas of the spectrum seem to be the most consistent. Once these points are located, the aerosol "pseudo-continuum" can be fit to the power law and the aerosol optical depth determined. At this point, the various components of the optical depth are determined and can be input into the radiative transfer models.

2.2 Reference Target Reflectance

The reference targets used for our model validation experiments are bodies of water, large enough to not exhibit shore and adjacency effects and deep enough so that the bottom does not reflect any measurable radiance into the sensor aperture. There are many such bodies of water in

the Canadian Shield. A set of measurements of the upwelling radiance from the surface and volume of Grand Lake, some 25km north of Ottawa, was made. The measurements were made at view angles ranging from 60 degrees to nadir at azimuth increments (relative to the sun) of 45 degrees. This measurement set permits a full angular characterisation of the reflectance of the water surface, necessary for off-nadir pointing sensors. Data were acquired on four different days throughout the summer observing season and, on one day, measurement sets were obtained throughout the day to include the effect of changing solar illumination angle. An example of surface and volume reflectance spectra can be found in Figures 3 and 4. These data were obtained with a Spectron SE-590 portable spectrograph. Unfortunately, the absence of a spectral order separation filter, which allows blue light from the first order spectrum to fall on detectors in the red region of the zeroth order spectrum, contaminates the red signal, which is very small for water, and produces an erroneous reflectance spectrum. Useable data in this situation are limited to the spectral interval between 400nm and 800nm. This is not such a problem as the reflectance of water beyond 800nm is of the order of a few tenths of a percent. Future measurements will include filters, so that the near infrared region can be accurately determined as well as polarisers so that the highly polarising nature of water surfaces at certain angles can also be characterised.

2.3 PATH RADIANCE

Path radiances over the measured reference targets were obtained from SPOT-HRV and Landsat TM images, as well as from a downward looking SE-590 spectroradiometer which was flown in a Falcon fan jet at several different altitudes. A measurement set similar to the water reflectance was obtained

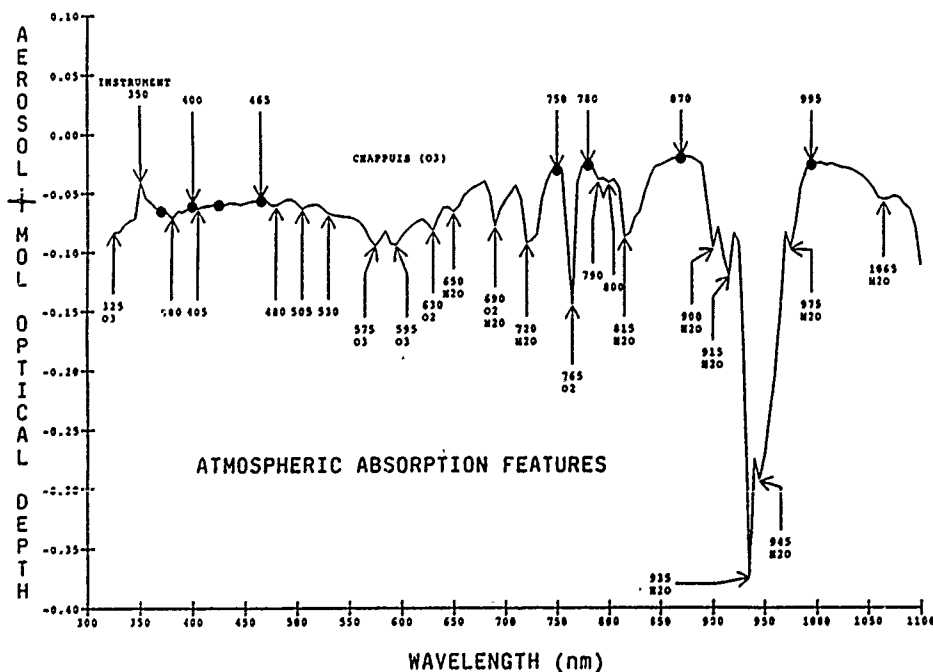


Figure 2:

Rayleigh subtracted optical depth spectrum. Optical depth changes are exhibited by aerosols and molecules. Some of the more prominent molecular absorptions are identified. The large black dots represent typical "pseudo-continuum" points.

with the SE-590, that is at view angles ranging from 60 degrees through nadir at relative azimuth increments of 45 degrees. This data set will characterise the off-nadir aspects of the path radiance for validation by a radiative transfer code. Also, some measurement sets were made in polarised light, adding this characterisation to the model validation. This work is currently in progress.

GRAND LAKE REFLECTANCE

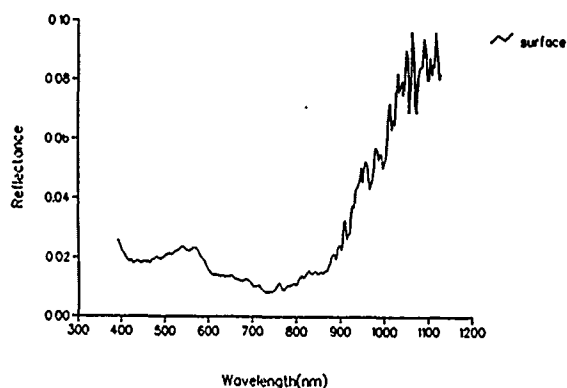


Figure 3: Surface reflectance of Grand Lake (at nadir). The water contained pigments giving rise to the familiar "green bump" as seen in vegetation spectra.

GRAND LAKE REFLECTANCE

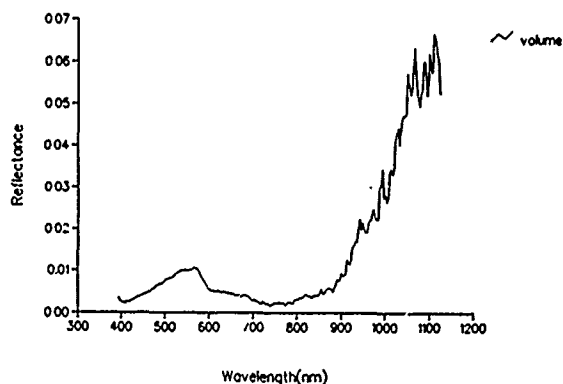


Figure 4: Volume (subsurface) reflectance of Grand Lake (at nadir) which is spectrally similar to the surface reflectance but significantly lower.

3. CONCLUDING REMARKS

The conduct of a measurement program for the validation of radiative transfer models over two observing seasons has revealed some notable points of interest. An important aspect in the determination of optical depths is the determination of the exo-atmospheric instrument response. Unless the solar observations are made at a site which is free of aerosols (usually high altitude like Mauna Kea, Hawaii), several days of observations must be made in relatively aerosol free conditions in order to obtain an accurate calibration. This usually means that a substantial commitment must be made to obtain observations over periods of several months at sites of low elevation. The accuracy of the optical depth determinations depends not only on the care with which the solar radiometry is made but also on the care with which the Rayleigh optical depths are computed and the "continuum" points are chosen for the calculation of the aerosol optical depths. The value of high spectral resolution optical depth data has been demonstrated in that it permits the careful investigation of the above-mentioned problems, as well as allowing correlations to be made with visual sky and meteorological parameters at the time of the observations, work which is currently in progress.

4. REFERENCES

1. Reagan, J.A., Scott-Fleming, I.C., Herman, B.M. and Schotland, R.M., "Recovery of spectral optical depth and zero-airmass solar spectral irradiance under conditions of temporally varying optical depth", Proceedings of IGARSS '84, Strasbourg, France, August 1984.
2. Reagan, J.A., Thomason, L.W., Herman, B.M. and Palmer, J.M., "Assessment of atmospheric limitations on the determination of the solar spectral constant from ground-based spectroradiometer measurements", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-24, No. 2, pp 258-266, 1986.
3. Teillet, P.M., "Rayleigh optical depth comparisons from various sources", Private Communication, 1989.

AEROSOL ANOMALIES IN NIMBUS-7 COASTAL ZONE COLOR SCANNER DATA OBTAINED IN JAPAN AREA

Hajime Fukushima*, Raymond C. Smith**, Yasuhiro Sugimori*,
Mitsuhiro Toratani*, and Yoshizumi Yasuda***

- * School of Marine Science and Technology, Tokai University, Orido, Shimizu,
424 Japan.
** Center for Remote Sensing and Environmental Optics, University of
California at Santa Barbara, Santa Barbara, California, 93106 U.S.A.
*** Faculty of Engineering, Chiba University, Yayoi-cho, Chiba, 260 Japan

ABSTRACT

About 400 CZCS scenes covering the Japan area in November 1978 - May 1982 were processed to study the applicability of the Gordon-Clark atmospheric correction scheme which produce water leaving radiances L_w at 443, 520 and 550 nm as well as phytoplankton pigment maps. Typical spring-fall aerosol radiance in the images were found to be 0.8 - 1.5 $\mu W/cm^2 \cdot nm \cdot sr$, which is about 50% more than reported for the U.S. eastern coastal images. The correction for about half the data resulted in negative $L_w(443)$ values, implying overestimation of aerosol effect for this channel. Several possible reasons for this are considered, including deviation of aerosol optical thickness τ_a at 443 nm from that estimated by the Angstrom's exponential law which the algorithm assumes. A routine observation of solar irradiance was conducted on the ground during late spring - summer in the year 1988 to observe the variability in aerosol optical thickness and the deviation of τ_a from the Angstrom's exponential law. The analysis shows that, assuming use of the Gordon-Clark algorithm, and for the pigment concentration of about $1 \mu g/l$, -40 to +100% error in satellite estimates is expected commonly. Although this does not fully explain the negative $L_w(443)$ in satellite data, it seems to contribute to the problem significantly together with other error sources including one in the sensor calibration.

1. Introduction

Nimbus-7 Coastal Zone Color Scanner (CZCS) is a scanning radiometer which views ocean at 4 visible wavelengths, 443, 520, 550 and 670 nm, to map the oceanic phytoplankton pigment field. NASA has started to process all the CZCS data to produce global pigment maps each averaged over every 2 weeks during the CZCS life time (1978 to 1986). The Gordon-Clark atmospheric correction scheme was adopted as their standard processing procedure, which assumes that the total radiance L_t at wavelength λ observed by the satellite sensor is decomposed as

$$L_t(\lambda) = L_m(\lambda) + L_a(\lambda) + t(\lambda)L_w(\lambda), \quad (1)$$

where L_m is the contribution arising from aerosol scattering, and $t(\lambda)L_w(\lambda)$ is the water-leaving radiance diffusely transmitted towards the satel-

lite. According to their algorithm, L_m is calculated based either on the single scattering approximation (Gordon, 1983) or on the multiple scattering radiative transfer model (Gordon, 1988) which include polarization effect but not interaction with aerosol particles. t is also calculated assuming no aerosol effect including absorption. The process of atmospheric correction proceeds pixel by pixel in the following manner. First, given $L_t(670)$, and putting $L_w(670)$ be zero, we get $L_a(670)$ from Eq. 1. Then assuming that $L_a(\lambda)$ is proportional to the product of $\tau_a(\lambda)$ and the extraterrestrial solar irradiance E_0 (as is shown in Eq. (4) later), and also assuming that the following Angstrom's exponential law holds,

$$\tau_a(\lambda) = \tau_a(670) \left(\frac{\lambda}{670} \right)^{-\alpha} \quad (2)$$

$L_a(\lambda)$ at $\lambda = 443, 520$ and 550 nm are obtained. Here, α is the Angstrom exponent which can be determined for each CZCS scene by the "clear water radiance concept" (Gordon, 1988) or some other procedure. Knowing $L_a(\lambda)$ gives $L_w(\lambda)$ values whose combination in turn gives the pigment concentration estimate for each pixel.

For the Japan area, CZCS left more than 400 scenes of worth analyzing with low enough cloud coverage. But little has been done on the processing of these local data partly because of lack of the confidence on the processing algorithms. Our preliminary analysis shows the Gordon-Clark algorithm sometimes resulted in rather unlikely pigment maps.

In this paper, we first shows results of atmospheric corrections conducted over CZCS Japanese scenes. Our aim is to check the soundness of the Gordon-Clark algorithm when applied to the area of relatively high humidity and under the influence of terrigenous aerosols from the mainland China. Secondly, we will see the expected error in the satellite estimates of phytoplankton pigment concentration based on the variability of aerosol optical thickness measured routinely from the ground in 1988.

2. CZCS Data Analysis

2.1 Data Set

Among all the CZCS data that cover the Japan area (20 - 40° N, 120 - 160° S) during the period

from November 1979 through May 1982, total of 410 scenes with low cloud coverage (1/4 or more cloud free area in the image) were chosen and subjected to study. Each scene was then atmospherically corrected using the Gordon-Clark algorithm (Gordon, 1983 and Gordon, 1988) to produce Lw(443) and La(670) images as well as a pigment map. In obtaining Lw(443), we set the Angstrom exponent to zero which is adopted widely in CZCS data processing as a typical value for marine type aerosol.

2.2 Seasonal Change in Aerosol Scatterance

All the color-scaled La(670) images were scanned by eye inspection and categorized into 3 groups in terms of mean La(670) level: (1) mean La < 0.7 $\mu\text{W}/\text{cm}^2 \cdot \text{nm} \cdot \text{sr}$ (clear image), (2) mean La between 0.7 - 1.0 (moderately hazy) and (3) mean La > 1.0 (very hazy). Here, the averaging was done by human eyes and probably has some ± 0.1 uncertainty. Fig. 1 shows the ratio of moderately/very hazy scenes against total for each season during the period. In the figure the seasons are abbreviated: "S" for spring (March - May) and for summer (June - August), "F" for fall (September - November), and "W" for winter (December - February). As it shows Japanese spring-summer CZCS scenes have relatively high La(670) level compared to those in the eastern U.S. coastal area where typical La(670) values of 0.5 - 1.0 $\mu\text{W}/\text{cm}^2 \cdot \text{nm} \cdot \text{sr}$ are reported (Gordon, 1987). This may imply some difficulty in applying the Gordon-Clark atmospheric correction which assume separability of the Rayleigh and aerosol components.

2.3 Overcorrection of the Atmospheric Effect

Atmospheric correction for some scenes resulted in areas with negative water leaving radiance at 443 (Lw(443)) due to overcorrection of aerosol scatterance. Subjective categorization of the scenes similar to that for La(670) (Fig. 1) was conducted with respect to the extent of the negative Lw(443) area. The categories are: (1) scenes where negative Lw(443) values are found, (2) scenes with more than half the cloud-free area covered with negative Lw(443), and (3) the rest. Those scenes covered widely with negative Lw(443) are typically accompanied by severe discontinuity in pigment map, or artificial chlorophyll front in the image. Fig. 2 shows the result in the same way as in Fig. 1. This negative Lw is considered to be caused by some non-oceanic aerosols since, judged from Lw(550) values, more likely Angstrom exponent for these scenes seems to have a higher value than the assumed (zero), indicating presence of some terrigenous aerosol. (If we were to take higher value as the exponent, more severe negative Lw(443) would result in.) The sensor degradation/calibration problem is also suspected since the figure shows rather weak correlation of the negativity with the La(670) variability shown in Fig. 1.

3. Variability of Aerosol Optical Thickness and its Implication on Satellite Estimates of Pigment Concentration

3.1 Ground-based Aerosol Optical Thickness Measurement

We measured aerosol optical thickness through spring-winter in 1988 at Shimizu, a Japanese city located on the Pacific coast of the central Japan.

The direct solar irradiance was measured at 13 wavelengths in 410 ~ 867 nm region by the Gershun tube method based on the WHO recommendation. Aerosol optical thickness $\tau_a(\lambda)$ for each wavelength λ was calculated by the Lambert-Bouguer law,

$$E_d = E_0 \exp(-(\tau_m + \tau_{oz} + \tau_a)/\cos \theta), \quad (3)$$

where E_0 is the extraterrestrial solar irradiance, E_d is the measured direct solar irradiance on the ground, τ_m is the Rayleigh (or molecular) optical thickness, τ_{oz} is ozone optical thickness, and θ is the solar zenith angle. The values $E_0(\lambda)$ were derived from Neckel and Labs(1981) whereas τ_m and τ_{oz} were obtained from LOWTRAN 6 code.

3.2 Deviation of measured $\tau_a(440)$ from the Angstrom's law

Fig.3 shows plots of the measured aerosol optical thicknesses at 440nm through the observation period. The data sparsity in summer is due to the anomalous cloudy/rainy weathers we had in the year. As the figure shows, τ_a value is generally high in spring-summer and is low in winter with much variability in fall.

Two examples of $\tau_a(\lambda)$ plots as a function of wavelength are shown in Fig.4. The solid lines give the least square error on the measured values at 515, 660, 770 and 867nm. Although the regression can be regarded as a good approximation to the Angstrom's exponential law, there usually is slight deviation of measured $\tau_a(440)$ from the value estimated by the least square error line, or the Angstrom's law. As shown in Fig.5, the deviation does not correlate with the magnitude of measured $\tau_a(440)$ with the standard deviation of 0.03.

3.3 Errors in Satellite Estimates of Pigment Concentration due to the deviation of Aerosol Optical Thickness at 440nm

Assuming a "Bio-Optical" algorithm and an aerosol scattering phase function, we can calculate the expected estimation error in satellite-derived pigment concentration based on the Gordon-Clark atmospheric correction scheme. First, we adopt a two-term Henyey Greenstein function as an aerosol phase function P_a in the following equation.

$$L_a(\lambda) = \frac{E_0(\lambda) \omega_a(\lambda) \tau_a(\lambda) P_a(\Psi)}{\cos \theta} \quad (4)$$

Here, E_0 is the extraterrestrial solar irradiance reduced by two trips through the ozone layer, ω_a is scattering albedo of aerosol particle (we assume ω_a is unity), Ψ is aerosol scattering angle, and θ is the solar zenith angle.

The equation provides the value of La(440) for a given τ_a value, thus giving $\Delta La(440)$, a simulated error in satellite estimate of the La due to the Angstrom's law. From Eq. (1), $\Delta Lw(440)$ is given by the relation,

$$\Delta Lw(440) = \Delta La(440) / t(440). \quad (5)$$

We can also calculate a nominal value of Lw(440) for a given pigment concentration C assuming a "Bio-optical" algorithm (Gordon, 1983),

$$C = 1.12 \left(\frac{Lw(550)}{Lw(440)} \right)^{1.7} \quad (6)$$

on the assumption that Lw(550) is known and calculated by

$$Lw(550) = nLw(550) \cdot \exp(-(\tau_w/2 + \tau_{oz})/\cos \theta) \cdot \cos \theta \quad (7)$$

where $nLw(550)$ is "normalized water leaving radiance" at 550 nm and is constant, $0.3 \mu W/cm^2 \cdot nm \cdot sr$. Combining (5) and (6) provides a final pigment estimation error

$$\Delta C = \frac{C' - C}{C} \quad (7)$$

where C' is derived from the biased $Lw(440)$.

Fig. 6 shows the final result of error analysis for 4 assumed pigment concentration levels. Recalling that the standard deviation of $\tau_a(440)$ in our observation was about 0.03, the figure implies $\pm 20\%$ (at $C=0.2 \mu g/l$) to $\pm 150\%$ (at $C=1.5 \mu g/l$) error in C will be fairly common.

4. Conclusion

The atmospheric correction for the CZCS Japan scenes produces rather unlikely $Lw(443)$ values for several possible reasons: (1) dense aerosol concentration, (2) influence of terrigenous aerosols as reported recently (Iwasaka, 1988), and (3) sensor calibration error. The first two aspects will be observed through the spectral behavior of aerosol optical thickness, or the deviation of $\tau_a(443)$ from the Angstrom's exponential law. Our analysis on the variability of measured $\tau_a(443)$ showed that the law occasionally reduces $Lw(443)$ to result in as much as $\pm 150\%$ error in satellite estimate of phytoplankton pigment. Although this does not fully explain the anomalous $Lw(443)$ values in the satellite data, the authors consider that this contributes significantly to the problem together with other error sources including one in the sensor calibration.

The results of our study suggest modification of

the atmospheric correction algorithm to include the interaction in aerosol-molecule multiple scattering as well as a new scheme estimating $\tau_a(440)$ independently of the Angstrom's law.

The image analysis in this study was done on the University of California Marine Bio-Optics (UCMBO) image processing system at the University of California, Santa Barbara under the support of the National Aeronautics and Space Administration (NASA) Grant No. NAGW-290-3.

REFERENCES

1. Gordon, H.R., D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, and W.W. Broenkow, "Phytoplankton pigment concentration in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates", *Applied Optics*, 22, 20, 1983.
2. Gordon, H.R., and D.J. Castano, "Coastal Zone Color Scanner atmospheric correction algorithm: multiple scattering algorithm", *Applied Optics*, 26, 11, 1987.
3. Gordon, H.R., J.W. Brown, and R.H. Evans, "Exact Raleigh scattering calculations for use with the Nimbus-7 Coastal Zone Color Scanner", *Applied Optics*, 27, 5, 1988.
4. Viollier M., D. Tanre and P. R. Deschamps, "An algorithm for remote sensing of water color from space", *Boundary-Layer Meteorology*, 18, pp247-267, 1980.
5. Neckel, H., and D. Labs, "Improved data of solar spectral irradiance from 0.33 to 1.25 μ ", *Solar physics*, 74, pp231- , 1981.
6. Iwasaka, Y., M. Yamamoto, R. Imasu and A. Ono, "Transport of Asian dust (KOSA) particles: Importance of weak KOSA events on the geochemical cycle of soil particles", *Tellus*, 40B, pp494-503, 1988.

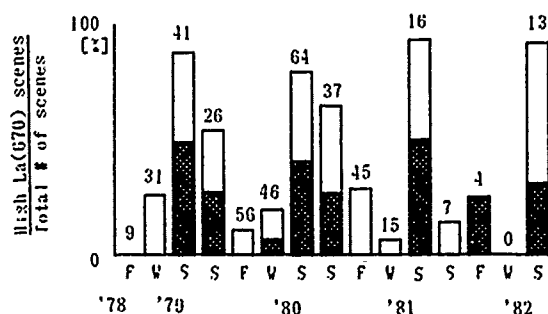


Fig. 1 Seasonal change in the ratio of high aerosol scenes in the Japanese CZCS data set. Total height of each bar represents those scenes with rough averaged $La(670)$ more than $0.7 \mu W/cm^2 \cdot nm \cdot sr$ whereas hatched section corresponds to the scenes with average $La(670)$ over $1.0 \mu W/cm^2 \cdot nm \cdot sr$. The numbers shown indicate total number of analyzed scenes for each season.

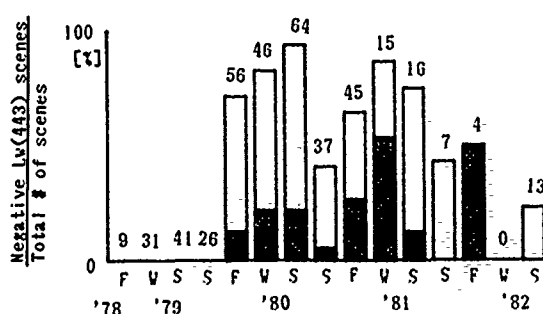


Fig. 2 Seasonal change in the ratio of CZCS scenes that result in negative channel 1 value after atmospheric correction. Hatched bar corresponds to the scenes that are widely (more than half the cloud-free area) covered with negative $Lw(443)$. Total height of bar corresponds to those scenes that have negative $Lw(443)$. The numbers shown indicate total number of analyzed scenes for each season.

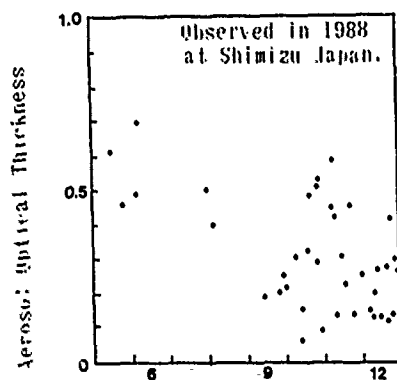


Fig. 3 Monthly change in Measured aerosol optical thickness at 440nm.

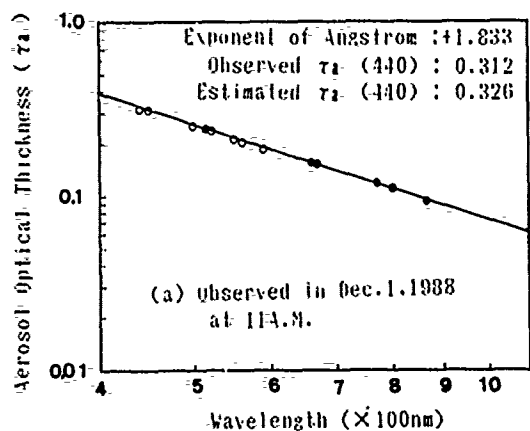


Fig. 4 (a) Examples of Measured Aerosol Optical Thicknesses. Solid line shows regression on the measurements at 515, 660, 770 and 867nm (*).

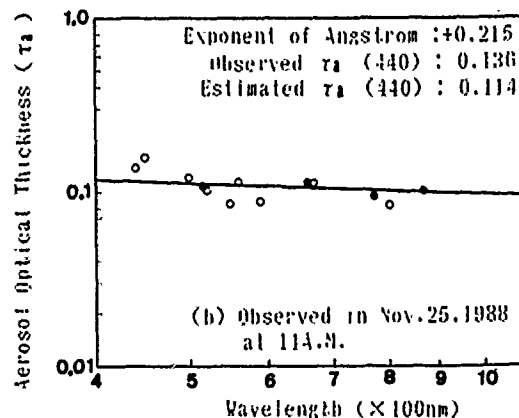


Fig. 4 (b) Examples of Measured Aerosol Optical Thicknesses. Solid line shows regression on the measurements at 515, 660, 770 and 867nm (*).

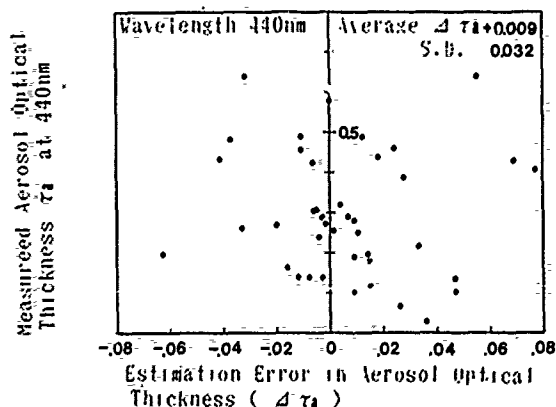


Fig. 5 Difference between measured and estimated aerosol optical thicknesses at 440nm. Estimation was made via the Angstrom's exponential law.

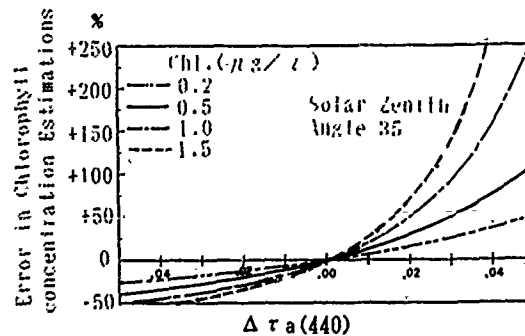


Fig. 6 Estimation error in Chlorophyll concentration due to erroneous aerosol optical thickness values at 440 nm.

The Distribution of Clear-Sky Radiation Over Varying Terrain

RALPH DUBAYAH, JEFF DOZIER AND FRANK DAVIS

*Department of Geography and Center for Remote Sensing and Environmental Optics,
University of California, Santa Barbara*

ABSTRACT

This research explores the topographic distribution of clear-sky incoming solar radiation over varying terrain. Using a two-stream atmospheric radiation model and digital elevation models (DEM) of several different physiographic regions, we simulate clear-sky radiation at 15 times throughout the day for three dates, December 15, March 15 and June 15. Geostatistical analysis is used to characterize the spatio-temporal variability in modeled radiation over each terrain grid. This analysis reveals that, for any particular elevation model and set of atmospheric conditions, the variance and spatial autocorrelation of the data depend on sun angle and on probability density functions of slope. By considering the behavior of the radiation mean and variance as a function of sun angle and slope, we derive predictive equations for these statistics. Mean irradiance is a function of slope distribution, but can be approximated as function of mean slope only. The variance in irradiance is approximately proportional to $\sin^2 \bar{S}$, where \bar{S} is the mean slope across the region. In addition, there is one particular solar zenith angle at which variance is maximized, and this is the same for all landscapes. This angle is a function of optical depth only and independent of any particular terrain considerations.

INTRODUCTION

This research explores the topographic distribution of solar radiation over varying landscapes. Topographically induced radiation variation affects many physical processes — examples are biomass distribution, evapotranspiration, snow accumulation and runoff — yet its spatial variability through time at regional levels is poorly understood. Solar radiation is also a key parameter in regional and global climate models, which increasingly use satellite based observations of radiation as inputs. Characterization of local spatio-temporal variation in radiation is fundamental to understanding these remotely sensed data, because variation exists at scales above and below sensor scales.

Our first objective is to model daily and seasonal variation in instantaneous incoming solar radiation, over wavelengths from 0.28 to 2.8 μm , using digital elevation data for each terrain. We then characterize the spatio-temporal variability of this radiation through geostatistical analysis. Finally, since radiation modeling is an expensive process, we compare the radiation regimes of different terrains by relating radiation variability, specifically the mean, variance and spatial auto-correlation, to topographic variability within each terrain. By doing so, we hope to find a method of estimating radiation variability solely through terrain considerations.

BACKGROUND

Seasonal and latitudinal variations in incoming radiation are well described, but only a few investigators have analyzed topographically induced variability in daily and seasonal radiation patterns across a region (e.g. Williams et al., 1972; Holland and Steyn, 1975; Dozier and Outcalt, 1979; Gates, 1980; Dozier, 1980; Kirkpatrick and Nunez, 1980; Davis, et al., 1988). We have previously examined topographic variation across the Konza Prairie, Kansas using DEM's of differing resolution as part of the First ISLSCP Field Experiment (FIFE) (Davis, et al, 1988; Dubayah, et al, 1989). For the Konza we found that the variance and spatial autocorrelation in clear-sky radiation depended on the resolution of the digital elevation grid and on sun angle. As grid resolution increased, variance increased and spatial autocorrelation decreased, reflecting a change in slope distribution. The autocorrelation also increased slightly with small and large zenith sun angle. This work led us to explore terrain variability in the more general framework provided here.

In the solar spectrum, slopes are irradiated from three sources: (1) F_s , direct irradiance from the sun; (2) F_d , diffuse irradiance from the sky, where a portion of the overlying hemisphere is obscured by terrain; and (3) F_r , direct and diffuse irradiance reflected from nearby terrain. The necessary calculations for obtaining these three quantities within the context of a two-stream atmospheric radiation model are given by Dozier (1988). The direct flux on a slope is

$$F_s = S_0 e^{-\tau_0/\cos\theta_0} \cos(i) \quad (1)$$

where S_0 is the exoatmospheric irradiance and τ_0 is the optical depth. The cosine of the illumination angle is given by:

$$\cos(i) = \cos\theta_0 \cos S + \sin\theta_0 \sin S \cos(\phi_0 - A)$$

where ϕ_0 and θ_0 are the solar azimuth and zenith angles, and S and A are the slope and azimuth of the slope.

Most of the variability in incoming clear-sky radiation is within the direct flux term and several useful results can be derived from it. The mean direct flux over the landscape is given by integrating over all slopes using independent probability density functions for slope and aspect, $\Phi(S)$ and $\Phi(A)$, of the particular terrain:

$$\bar{F} = S_0 e^{-\tau_0/\cos\theta_0} \int_{-\pi}^{\pi} \Phi(A) \int_0^{\pi/2} \Phi(S) \cos(i) dS dA \quad (2)$$

If the distribution of aspects is uniform, then $\Phi(A) = 1/2\pi$ and the mean flux is

$$\bar{F} = S_0 e^{-\tau_0/\cos\theta_0} \cos\theta_0 \int_0^{\pi/2} \Phi(S) \cos S dS \quad (3)$$

In the discrete case of calculating the mean over a digital elevation grid, we sum over all slopes using the slope frequency distribution.

We similarly form an equation for the direct flux variance as

$$\int_{-\pi}^{\pi} \Phi(A) \int_0^{\pi/2} \Phi(S) (F_s - \bar{F})^2 dS dA \quad (4)$$

Let us reconsider the equations for mean and variance assuming a uniform distribution for aspect and a constant mean slope, \bar{S} . The direct mean flux is

$$\bar{F} = S_0 e^{-\tau_0/\cos\theta_0} \cos\theta_0 \cos\bar{S} \quad (5)$$

and the variance over a landscape with these properties is:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} (F_s - \bar{F})^2 dA = \frac{1}{2} S_0^2 e^{-2\tau_0/\cos\theta_0} \sin^2 \bar{S} \sin^2 \theta_0 \quad (6)$$

Using Equations 3 and 5 we can compare the effects of slope distribution and average slope on radiation variability between landscapes. Again, our assumption here is that for clear-sky conditions, even in areas of high-relief, most of the spatial and temporal variability in total incoming radiation will be related to the direct flux. Except for local horizon effects and changes in optical depth due to elevation, the diffuse radiation between landscapes is similar, because we use the same atmospheric parameters throughout the modeling process.

Equation 6 shows that under our simplifying terrain assumptions, radiation variance is proportional to the average slope. Under identical atmospheric conditions, the relative variability between two landscapes is proportional to the ratio of their slopes, i.e.:

$$\text{var}(F_1) = \frac{\sin^2 S_1}{\sin^2 S_2} \text{var}(F_2) \quad (7)$$

where F_1 and F_2 is the flux on each landscape with mean slopes of S_1 and S_2 , respectively.

We also note that the solar angle at which maximum variance occurs across a landscape is obtained by taking the derivative of Equation 6 with respect to solar zenith angle and setting the result equal to 0. This leads to a unique maximum at a value of θ_0 such that

$$\frac{\cos^3 \theta_0}{\sin^2 \theta_0} - \tau_0 = 0 \quad (8)$$

In summary, under an assumption of uniform aspect, the mean and variance depend on the slope probability distribution function of the particular terrain. If we further generalize to a constant slope distribution, the magnitude of the variance depends on the average slope, but the sun angle at which the maximum variance occurs is independent of slope, being completely determined by the optical depth. For any given optical depth there is a solar zenith angle at which variance is maximized.

METHODOLOGY

In order to test the validity of the relationships derived above on varying terrain, we obtained digital elevation models (DEM's) of four different landscapes at 30 m resolution corresponding to U.S.G.S. 1:24,000 quadrangles. These were: Swede Creek, Kansas, a low prairie landscape; La Honda, California, from the northern coast range province; Los Machos, California, from the southern transverse range; and, Mt. Tom, California, a glacial landscape from the Sierra Nevada.

The mean slope and relative relief of each DEM are listed in Table 1. Slope histograms are shown in Figure 1.

TABLE 1. Digital elevation data.

| DEM | Type | Relief (m) | S(°) |
|-------------|------------------|------------|------|
| Swede Creek | prairie | 144 | 5.5 |
| La Honda | coast range | 620 | 16.3 |
| Los Machos | transverse range | 600 | 17.2 |
| Mt. Tom | alpine | 1550 | 24.6 |

Both the Mt. Tom and Los Machos data are the standard DEM product from the U.S.G.S. The Mt. Tom DEM suffers from systematic noise resulting from the digitization process. The La Honda and Swede Creek models are derived from direct digitization of contour lines from topographic maps. From each DEM we use a representative 10 km × 10 km region.

We simulated clear-sky solar radiation (0.28–2.8 μm) for December 15, March 15 and June 15 over each terrain using a two-stream atmospheric radiation model with an optical depth of 0.2, a substrate reflectance of 0.3, a single-scattering albedo of 0.85, and a scattering asymmetry parameter of 0.52. Model parameters were held constant for all landscapes at all times. On each date, 15 times, corresponding to the intervals for Kronrod quadrature between sunrise and sunset, were picked and the radiation over each of the terrains calculated. To characterize the spatio-temporal variability on each day, we calculated time-dependent semi-variograms using 4,000,000 randomly chosen pairs of points.

RESULTS

Table 2 gives a partial listing of irradiance means and variances before noon, sorted by solar zenith angle, for all three dates. Afternoon values are symmetric and azimuthal effects are small for these particular landscapes, although this is not true for landscapes in general.

TABLE 2. Radiation statistics.

| θ_0 (deg) | Swede | | La Honda | | Los Machos | | Mt. Tom | |
|---------------------|----------------------------------|---|----------------------------------|---|----------------------------------|---|----------------------------------|---|
| | \bar{F} (Wm ⁻²) | var (Wm ⁻²) ² | \bar{F} (Wm ⁻²) | var (Wm ⁻²) ² | \bar{F} (Wm ⁻²) | var (Wm ⁻²) ² | \bar{F} (Wm ⁻²) | var (Wm ⁻²) ² |
| 87 | 23 | 1 | 21 | 6 | 21 | 6 | 22 | 119 |
| 83 | 81 | 356 | 75 | 1250 | 80 | 1599 | 92 | 8020 |
| 80 | 138 | 1116 | 136 | 5132 | 128 | 5759 | 130 | 16164 |
| 78 | 189 | 1930 | 178 | 9114 | 138 | 10512 | 198 | 31516 |
| 72 | 318 | 3047 | 313 | 20496 | 301 | 23406 | 272 | 50520 |
| 69 | 366 | 3462 | 353 | 21866 | 340 | 25426 | 323 | 61980 |
| 66 | 452 | 3585 | 448 | 26088 | 443 | 29430 | 387 | 64540 |
| 61 | 532 | 4451 | 504 | 28733 | 522 | 32250 | 527 | 84994 |
| 55 | 630 | 3928 | 613 | 28151 | 599 | 31072 | 559 | 82970 |
| 51 | 739 | 4315 | 708 | 29207 | 725 | 29844 | 710 | 83103 |
| 42 | 888 | 3463 | 858 | 26384 | 871 | 23688 | 835 | 66966 |
| 38 | 944 | 2629 | 918 | 24380 | 924 | 20781 | 870 | 56973 |
| 23 | 1082 | 1421 | 1037 | 16343 | 1052 | 10314 | 1030 | 29148 |
| 13 | 1155 | 883 | 1114 | 12080 | 1122 | 4398 | 1080 | 16398 |

Fig. 2 shows time-dependent semi-variograms for June 15. The distance at which the semi-variance reaches its maximum or sill value on these plots is the range to which the data are spatially autocorrelated. The data show clear peaks in variance at mid-morning at the same time of day (solar azimuth) on each terrain, with strong decreases during high and low sun zenith angles. The range on all terrains increases with small/large zenith sun angles. This is expected

since at high zenith angles the direct component is small compared to the diffuse and diffuse radiation is isotropic in the two-stream model, leading to uniform illumination conditions. At low zenith angles, the differences in illumination due to topography also become less important leading to more uniform illumination and data which is correlated to longer distances. The average range for each terrain over all times is given in Table 3 along with semi-variogram ranges for slope and aspect. Ranges were calculated assuming an exponential

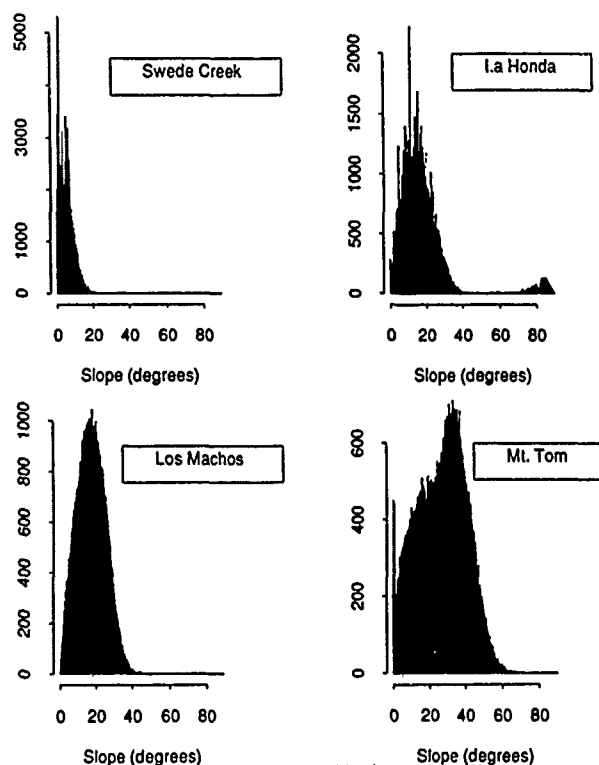


FIGURE 1

variogram model calculating a least-squares fit of the data to the model.

TABLE 3. Semi-variogram ranges

| DEM | F range (m) | Slope range (m) | Aspect range (m) |
|-------------|----------------|--------------------|---------------------|
| Swede Creek | 231 ± 28 | 275 | 233 |
| La Honda | 425 ± 35 | 363 | 416 |
| Los Machos | 367 ± 69 | 321 | 321 |
| Mt. Tom | 1202 ± 200 | 904 | 1073 |

DISCUSSION

Mean Radiation

We can use Equations 3 and 5 to examine how slope affects mean radiation over a region. Again assuming that the diffuse term will be relatively constant, Equation 3 predicts that the relationship between topographically modulated direct flux and the direct flux on a flat surface will be

$$F_S = F_{\text{flat}} \int_0^{\pi/2} \Phi(S) \cos S \, dS \quad (9)$$

Therefore the slope of a regression line between the flat flux and the flux over the terrain should be the integral in Equation 9. If we assume constant terrain slopes, the regression slope is simply $\cos S$. Fig 3 shows plots of the difference between the mean irradiance over a terrain (as calculated using the two-stream model and DEM data) and

the irradiance on a flat surface at the same sun position and atmospheric conditions. The relationships predicted using the mean terrain slope, S , the actual slope frequency distributions, and the regression line slope of the actual data are summarized in Table 4.

TABLE 4. Regression slopes for terrain irradiance vs. irradiance on flat surface.

| slope | Swede | La Honda | Los Machos | Mt. Tom |
|------------------------|-------|----------|------------|---------|
| data | 0.99 | 0.96 | 0.97 | 0.91 |
| $\cos S$ | 0.99 | 0.96 | 0.95 | 0.90 |
| $\int \Phi(S) \cos(S)$ | 0.99 | 0.95 | 0.95 | 0.88 |

Note that for a region with high average slopes, such as Mt. Tom, there is a 10% difference between the mean calculated over all slopes and that measured on a flat surface, say using a level pyranometer.

Radiation Variance

From Equation 8 we predict that the variance in irradiance should be maximized at a solar zenith angle of 58° for an optical depth of $\tau_0 = 0.2$, and that this angle should be same for all terrains. Figure 4 shows that this is indeed true. The patterns of variance as a function of zenith angle follow Equation 6, which uses only the mean slope, closely, and all landscapes show maximums near 58° .

We also predicted that the variances should be proportional to the ratio of the mean slopes between landscapes. For example, the S^*

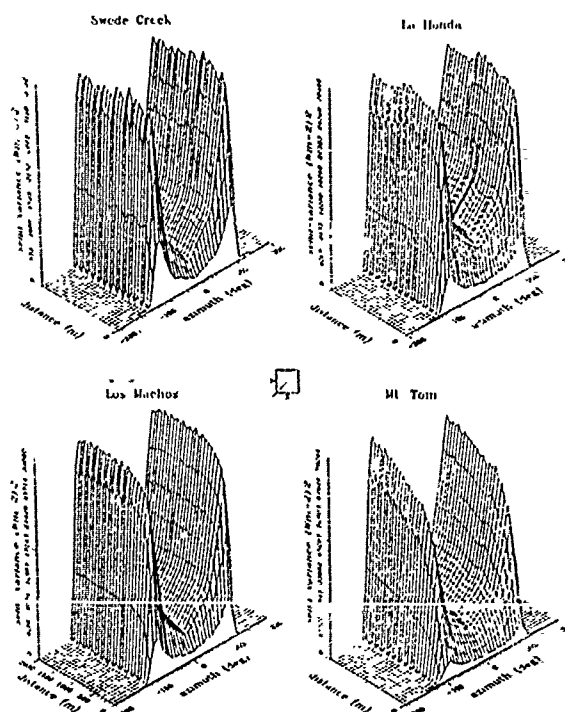


FIGURE 2

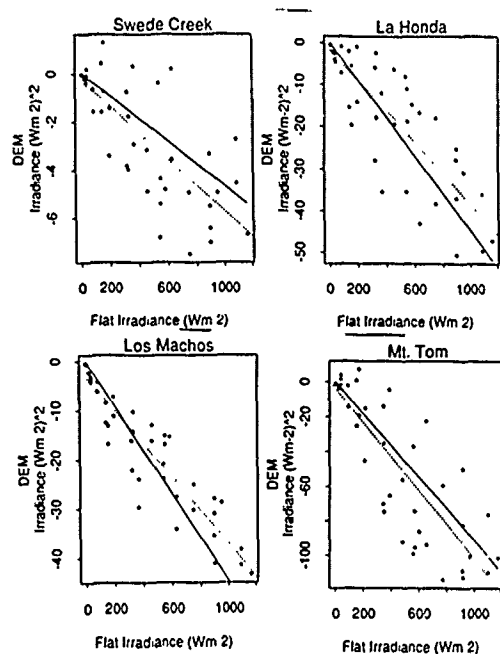


FIGURE 3

Creek radiation variance should be 0.053 of the Mt. Tom variance from:

$$\frac{\sin^2 \bar{S}_{\text{swede}}}{\sin^2 \bar{S}_{\text{tom}}} = \frac{\sin^2(5.5^\circ)}{\sin^2(24.6^\circ)} = 0.053$$

Table 5 gives predicted and actual variance factors for all terrain comparisons. The actual factors were obtained by regressing the radiation data between any two terrains. These relationships were always linear with an r^2 of over 0.97 significant at the 0.01 level.

TABLE 5. Predicted and actual variance factors.

| DEM | Mt. Tom b | Los Machos b | La Honda b |
|------------|--------------|-----------------|---------------|
| Swede | 0.05 | 0.05 | 0.11 |
| La Honda | 0.46 | 0.37 | 0.90 |
| Los Machos | 0.51 | 0.40 | 0.91 |

It appears we can approximately compare radiation variance among terrains by comparing their average slopes. We qualify this result, however, since for many terrains, the concept of a mean slope is vague, if not meaningless. For example, if the terrain covers many geomorphic types, if the model scale is quite small, or if the slope distribution is highly skewed, then the physical significance of an average slope is unclear. Also, if slopes are preferentially oriented, then the aspect distribution will significantly effect the mean and variance values, and convolve the relationship with mean slope. Nonetheless, within the limitations of the specific terrains and modeling scales chosen here, (i.e. 30 m to 10 km), radiation variability is clearly related to the average slope of these landscapes.

If the average slope controls the magnitude of the variation, what controls its spatial autocorrelation? This must be determined by the autocorrelation function of the cosine of the illumination angle, $\cos(i)$, which in turn is a function of sun angle and the autocorrelation functions of slope and aspect. Table 5 suggests that either the slope or aspect range is a good indication of the approximate autocorrelation to be expected in the radiation data. The overall exponential increase in radiation variance follows closely that of aspect, and to a lesser degree, slope.

CONCLUSION

We have modeled the topographic distribution of clear-sky incoming solar radiation using a two-stream atmospheric radiation model and digital elevation data for four different regions. Geostatistical analysis was used to characterize the spatio-temporal variability in modeled radiation over each terrain grid and revealed similar patterns in variance across each. For given sun angles and atmospheric conditions, the radiation means and variances were shown to be related to the probability density functions of terrain slope, and could be approximated by the mean terrain slopes. The relative means and variances among terrains could then be compared by considering linear functions of mean slopes. Finally, the spatial autocorrelation in radiation data is closely connected with the autocorrelation in slope and aspect.

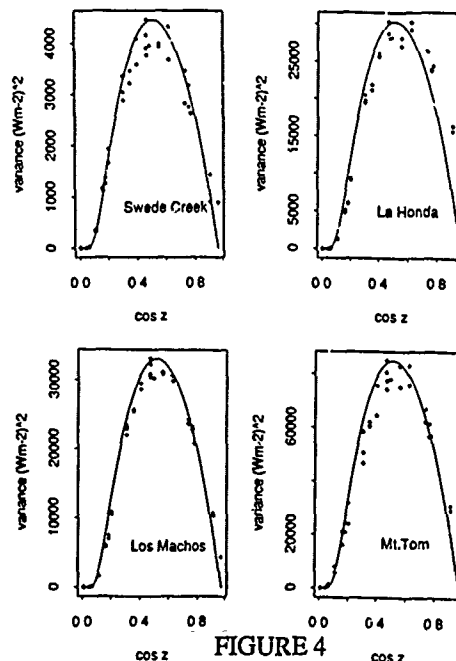


FIGURE 4

REFERENCES

- Davis, F., J. Dozier, and R. Dubayah, The topographic variation in solar radiation over the Konza Prairie, *Eos, Trans. Amer. Geophys. Union*, 68, 1988.
- Dozier, J., A clear-sky spectral solar radiation model for snow-covered mountainous terrain, *Water Resour. Res.*, 16, 709-718, 1980.
- Dozier, J., *Eos, Trans. Amer. Geophys. Union*, 69, 1988.
- Dozier, J. and S. I. Outcalt, An approach toward energy balance simulation over rugged terrain, *Geogr. Anal.*, 11, 65-85, 1979.
- Dubayah, R., J. Dozier, and F. Davis, *Topographic distribution of clear-sky radiation over the Konza Prairie, Kansas, USA*, Submitted to Water Resources Research, 1989.
- Gates, D.M., *Biophysical Ecology*, Springer-Verlag, New York, 1980.
- Holland, P.G. and D.G. Steyn, Vegetational responses to latitudinal variations in slope angle and aspect, *J. Biogeogr.*, 2, 179-183, 1975.
- Kirkpatrick, J.B. and M. Nunez, Vegetation-radiation relationships in mountainous terrain: eucalypt-dominated vegetation in the Ridson Hills, Tasmania, *J. Biogeogr.*, 7, 197-208, 1980.
- Williams, L.D., R.G. Barry, and J.T. Andrews, Application of computed global radiation for areas of high relief, *J. Appl. Meteorol.*, 11, 526-533, 1972.

EVALUATION OF ATMOSPHERIC EFFECT ON REMOTELY SENSED DATA BASED ON MEASUREMENT OF SPECTRAL RADIANCE

Y Yasuoka, T Miyazaki and Y Iikura

National Institute for Environmental Studies

16-2 Onogawa, Tsukuba, Ibaraki 305, Japan

Tel.: 0298-51-6111, Fax: 0298-51-4732

ABSTRACT

Atmospheric parameters were evaluated from the measurement of the upward surface radiance and the downward solar radiance above the surface in water area. The atmospheric transmittance and the path radiance were estimated by regression analysis between a set of radiance data which are the radiance detected by a remote sensor and the upward radiance simultaneously measured at the surface. Also the transmittance was estimated from the direct solar beam radiance measured at the surface by assuming the extraterrestrial solar radiance.

Key words: atmospheric effect, transmittance, path radiance, spectral measurement

1. INTRODUCTION

The atmospheric scattering and absorption affect the remotely sensed data and cause the serious classification errors or the estimation errors in analyzing data to get the various kinds of thematic maps. For example, in water area, the radiance from water surface holds only 10 to 30 % of the total radiance detected by the sensor on the satellite because of the low spectral reflectance of water bodies. Therefore the mapping of water quality parameters requires precise estimation of atmospheric effect and its removal.

There have been several methods to evaluate the atmospheric effect on the remotely sensed data (Otterman, 1980; Kaufman, 1988). In this paper, two methods are developed to quantitatively evaluate the atmospheric effect on the satellite data based on the field measurement of the spectral radiance above the water surface.

In the first method, the atmospheric transmittance and the path radiance are estimated by a linear regression analysis for a set of the radiance data which are the upward surface radiances measured at the stations and the remotely sensed radiances calculated from the CCT counts at the corresponding points in the multi-spectral images. The simple

radiative transfer equation shows that the slope and the y-intersect of the regression line give the estimates for the transmittance and the path radiance.

In the second method, first, the transmittance is estimated from the direct solar beam radiance measured at the surface by assuming the total solar radiance out of the atmosphere. Next the path radiance is estimated from the remotely sensed radiance and the upward surface radiance at the corresponding point using the estimated transmittance.

Spectral radiance measurements were carried out at Lake Kasumigaura and at Lake Biwa in Japan by utilizing a portable spectrometer during the overflight of LANDSAT and MOS-1. Both of the developed methods were applied to the LANDSAT MSS, TM data and the MOS-1 MESSR data to evaluate the atmospheric effect on them.

2. ATMOSPHERIC MODEL

Let $L(x)$ and $Q(x)$ be the upward radiance and the downward solar beam radiance at the altitude x (Fig.1). The radiance detected by a remote sensor is given by

$$L(H) = \tau \cdot L(0) + P \quad \dots (1)$$

where τ and P are the atmospheric transmittance and the path radiance between the earth surface and the remote sensor. The upward (nadir) radiance at the surface $L(0)$ is also given by

$$L(0) = a_0 \mu Q(0) + Q_s \quad \dots (2)$$

$$Q(0) = Q(H) \tau^{\downarrow} \quad \dots (3)$$

where a_0 is the earth surface reflectance, μ is cosine of the solar zenith angle and Q_s is the sky radiance at the surface. Also $Q(H)$ is the extraterrestrial solar radiance.

3. ESTIMATION OF ATMOSPHERIC TRANSMITTANCE AND PATH RADIANCE BY REGRESSION ANALYSIS

Let L_s and L_o be the remotely sensed radiance

and the simultaneously measured surface radiance at the i -th observation point. From the regression analysis between $(L(H))$ and $(L(O))$, we can get the regression equation

$$L(H) = \hat{a} L(O) + \hat{b} \quad \dots (4)$$

From the equation (1), it is easily shown that the regression coefficient \hat{a} and \hat{b} in equation (4) give the estimates for τ and P , respectively.

Figure 2 shows the examples of the regression line between the LANDSAT radiance and the upward radiance observed at Lake Kasumigaura on Nov. 24, 1981 and on March 3, 1982. Table 1 summarizes the estimated atmospheric transmittance (τ) and the path radiance (P) at Lake Biwa and at Lake Kasumigaura. In Fig. 3, the ratio (percentage) of the path radiance (P) in the total remotely sensed radiance ($L(H)$) over the water surface is illustrated for each band.

In the experiments at Lake Biwa and Lake Kasumigaura, the surface radiances were measured by a portable spectrometer whose spectral resolution was 2nm between 400 nm and 800 nm (Miyazaki, 1987). The original radiance was integrated over the spectral range of the LANDSAT MSS.

4. ESTIMATION OF ATMOSPHERIC TRANSMITTANCE AND PATH RADIANCE FROM THE MEASUREMENT OF SOLAR BEAM RADIANCE

From the equation (3), the atmospheric transmittance τ is given by

$$\tau = (Q(O)/Q(H))^\mu \quad \dots (5)$$

The equation (5) indicates that we can estimate the transmittance by measuring the direct solar beam radiance $Q(O)$ at the surface and also by assuming the extraterrestrial solar radiance $Q(H)$.

In this experiment, the solar beam radiance $Q(O)$ was indirectly estimated from the radiance of the white reflecting board (W) and from the radiance of the shadow on it (S) where the direct solar beam was cut off. The white reflecting board was made of Eastman Kodak white coating whose reflectance was almost 1.0 (Grum & Luckey, 1968). In the equation (2), using $a_o \approx 1.0$, $L(O) = W$ and $Q_s = S$, the direct solar beam radiance $Q(O)$ is given by

$$Q(O) = (W - S) / \mu \quad \dots (6)$$

The path radiance P was estimated from the equation (1) using the estimated transmittance τ , the upward radiance measured above the water surface $L(O)$ and the corresponding remotely sensed radiance $L(H)$ as follows:

$$P = L(H) - L(O) \cdot \tau \quad \dots (7)$$

Figure 4 shows the spectral characteristic of the atmospheric transmittance estimated from the observed radiance at Lake Kasumigaura on 7th November 1988 (8:15, 9:15, 10:15). Temporal change of the estimated transmittance is also illustrated in

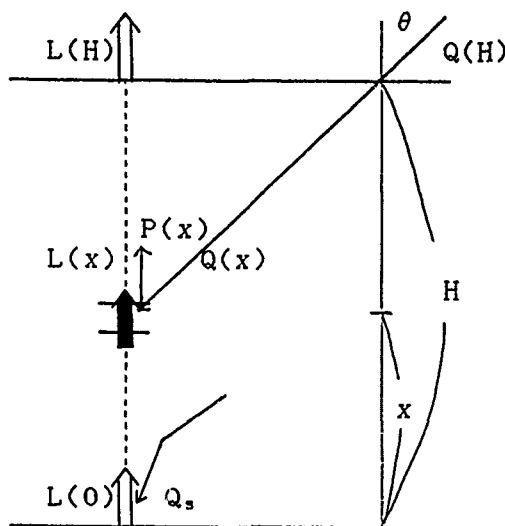


Fig.1 Optical process in the atmosphere.

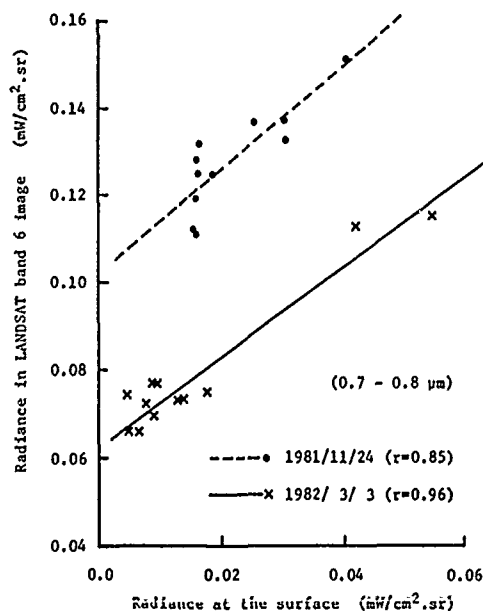


Fig.2 Regression analysis between the remotely sensed radiance (LANDSAT band 6) and observed surface radiance (Lake Kasumigaura; 1981/11/24 and 1982/3/3).

Table 1 Estimated atmospheric transmittance and path radiance
(Lake Biwa, 1985/10/3; Lake Kasumigaura, 1981/11/24, 1982/3/3).

| site | Lake Biwa | | Lake Kasumigaura | | | |
|------------------------------|--------------------------------------|------|--------------------------------------|------|--------------------------------------|------|
| date | 1985/10/3 | | 1981/11/24 | | 1982/3/3 | |
| band | τ $P(\text{mw/cm}^2/\text{sr})$ | | τ $P(\text{mw/cm}^2/\text{sr})$ | | τ $P(\text{mw/cm}^2/\text{sr})$ | |
| 4 (0.5 ~ 0.6 μm) | 0.79 | 0.22 | 0.77 | 0.26 | 0.86 | 0.26 |
| 5 (0.6 ~ 0.7 μm) | 0.72 | 0.13 | 0.72 | 0.16 | 0.78 | 0.11 |
| 6 (0.7 ~ 0.8 μm) | 0.94 | 0.09 | 1.17 | 0.10 | 1.01 | 0.06 |
| LANDSAT | 4 | | 2 | | 3 | |

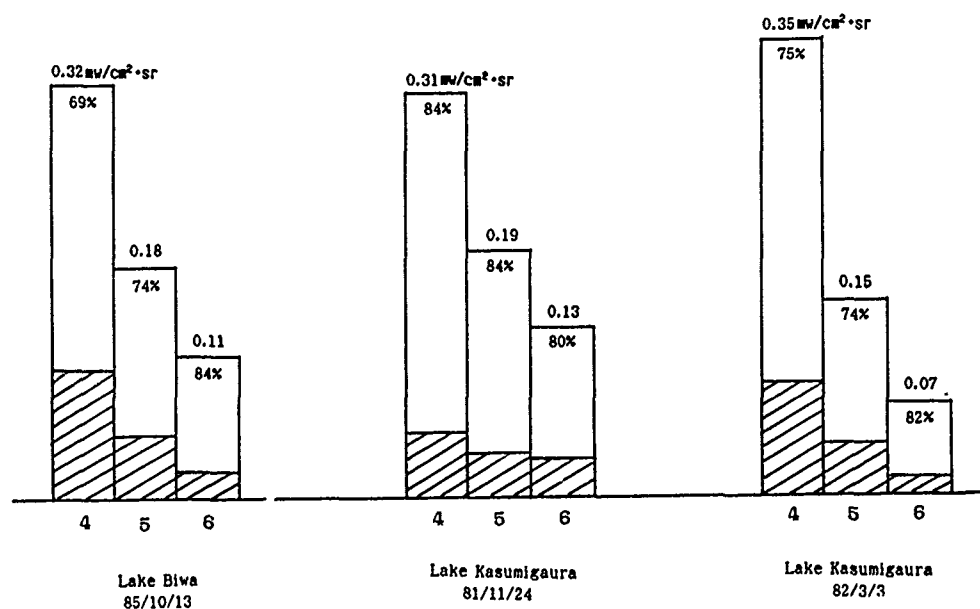


Fig.3 Ratio (%) of the estimated path radiance in the total remotely sensed radiance over the lake water.

Fig.5. Spectral radiance was measured by a portable spectrometer with 17 channels between 400 and 1050 nm.

5. CONCLUSIONS

In order to evaluate the atmospheric effect on the remotely sensed data, the atmospheric transmittance and the path radiance were estimated by two different methods based on the field measurement of spectral radiance. The methods were applied to Lake Biwa and Lake Kasumigaura to evaluate the atmospheric effect on remotely sensed data over the water surface. It

was shown that the path radiance over the lakes held 70 ~ 90 % in the total radiance detected by the remote sensor.

The atmospheric transmittance estimated at Lake Kasumigaura was compared with the transmittance estimated from the simultaneous measurement of aerosol distribution by the Laser Radar. The values from both methods showed quite good correspondence. The result indicates that the atmospheric parameters can be estimated in quite good accuracy by these simple field measurements of the spectral radiances.

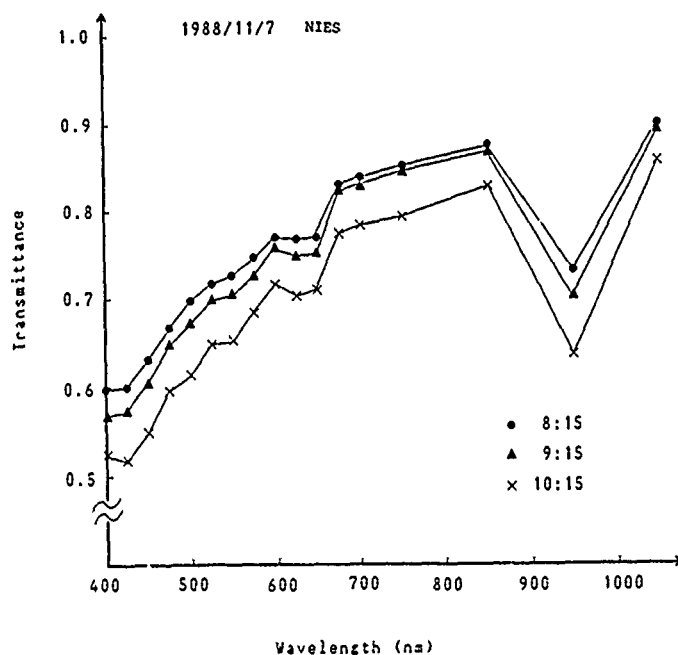


Fig.4 Spectral characteristics of the atmospheric transmittance (Lake Kasumigaura, 1988/11/7).

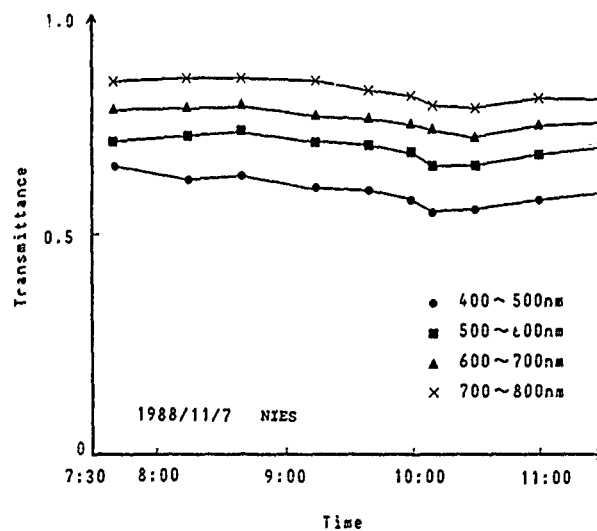


Fig.5 Temporal change of the atmospheric transmittance (Lake Kasumigaura, 1988/11/7).

REFERENCES

1. J. Otterman, "Atmospheric effect on radiometric imaging from satellites under low optical thickness conditions", *Remote Sensing of Environment*, Vol.9, pp115-129, 1980.
2. Y.J. Kaufman, "Atmospheric effect on spectral signature - Measurements and corrections", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.26, No.4, pp441-450, 1988.
3. T. Miyazaki, H. Shimizu and Y. Yasuoka, "High-speed spectroradiometer for remote sensing", *Applied Optics*, Vol.26, pp4761-4766, 1987.
4. F. Grum and G.W. Luckey, "Optical sphere paint and a working standard of reflectance", *Applied Optics*, Vol.7, pp2289-2294, 1968.

INTEGRATED ATMOSPHERIC CORRECTION OF MULTISPECTRAL SATELLITE IMAGERY

B. C. Robertson and B. Sharpe

MacDonald Dettwiler & Associates Ltd.
13800 Commerce Parkway, Richmond, B.C.
Canada V6V 2J3

P. M. Teillet

Canada Centre for Remote Sensing
2464 Sheffield Road, Ottawa, Ont.
Canada K1A 0Y7

ABSTRACT

Over the recent years, there has been an increasing trend towards the quantitative analysis of remotely sensed data. To meet this demand, an automated atmospheric correction system has been developed capable of reliably retrieving surface reflectance spectra from multispectral satellite data. An important phase of this program has been assessing the potential impact of the correction procedures on user applications by measuring the absolute and relative accuracies of the system. The results of this study—presented here—indicate that under reasonable conditions, surface reflectance spectra can be routinely retrieved to within 0.01 reflectance units in a production atmospheric correction system.

Keywords: Atmospheric Effects, Topographic Effects, Radiometric Correction.

1 INTRODUCTION

Quantitative analysis of multispectral satellite imagery such as that acquired by the Landsat TM and Spot HRV sensors depends on the ability to reliably extract surface reflectance properties from the satellite observed digital signal levels (DSL). This capability is hampered by the sensor-specific conversion of observed radiance to DSLs, and more importantly by the environmental perturbations introduced by the conditions under which the imagery was acquired. Sensor-specific characteristics are well understood and production systems are routinely retrieving the radiance to within 10% [Ahern 87]. Environmental perturbations, on the other hand, are significantly more difficult to characterize as they depend on temporally and spatially varying factors such as the illumination and observation geometries, and the atmospheric and topographic conditions under which the imagery was acquired [Teillet 86]. It is only recently that these radiometric errors have been corrected in production systems.

In areas of low relief, perturbations induced by solar and atmospheric conditions are the main source of radiometric errors often comprising over 80% of the total observed radiance over low reflectance targets in the shorter wavelengths. Over rugged terrain, radiometric errors caused by variations in the orientation of the surface with respect to the sun and the sensor typically dominate the signal. With

the increased use of multitemporal and multisensor datasets in applications such as thematic classification, change detection, and automated DEM extraction, it has become increasingly important to compensate for these errors. Further, the development of automated surface identification capabilities together with integrated spectral reflectance databases for different surface covers necessitates the absolute calibration of the raw satellite data to allow the extraction of quantitative, physically meaningful measurements.

Towards this goal, a prototype atmospheric correction package has been developed capable of reliably extracting surface reflectance spectra from the satellite observed DSLs. A study of the accuracy of the system, the results of which are reported here, assessed both the absolute accuracy by comparing reflectance derived from the imagery with coincident surface measurements, and the relative accuracy by assessing the ability of the system to extend spectral signatures of various surface covers between multitemporal datasets.

2 RADIOMETRIC ERRORS

For most remote sensing applications, the information of interest is the reflectance characteristics of the surface in the different regions of the electromagnetic spectrum. This information is perturbed by radiometric errors introduced into the data acquisition process both by sensor-specific characteristics and by the environment.

Remote sensing satellites measure the total radiance reflected from a particular area of the earth and convert this value to a DSL. The conversion from radiance to DSLs is a time-varying function of the physics of the sensor, thus retrieving radiance is a non-trivial process. Typically, inflight calibration data are used together with preflight measurements to calibrate the sensor. Recent work at the White Sands test site by Slater *et al* [Slater 87] has shown that, for the Landsat TM sensor, absolute calibration uncertainties as low as 3% can be achieved,

Having absolutely calibrated the raw satellite data, one may calculate the apparent reflectance of the surface as a simple ratio of observed radiance and the incident solar irradiance. Unfortunately, there is no simple equation relating the apparent reflectance to actual surface re-

flectance because of spatial and temporal variations in environmental conditions which introduce perturbations in both the radiation incident on and reflected from the earth's surface.

Environmentally induced errors take a number of forms. Radiation transmitted through the atmosphere is attenuated through the processes of molecular and aerosol scattering and absorption leading to a reduction in both the direct solar irradiance reaching the surface, and the surface reflected radiance reaching the sensor. The signal is further perturbed by additive noise introduced by radiation backscattered by the atmosphere (*path radiance*) as well as radiation scattered into the field of view of the sensor from surrounding surfaces (the so-called *adjacency effect*) [Tanré 81]. The net result of these effects is that dark targets appear brighter (due mainly to the additive noise) while bright targets appear darker (because of signal attenuation).

In areas of rugged terrain, the situation is further complicated by errors induced by the topography. Here, both the first order effects of local solar incidence angle and cast shadows, and the second order effects of variations in sky irradiance due to occlusion of the sky and radiation reflected off adjacent terrain are important [Woodham 87, Proy 86]. The situation is further complicated by the non-Lambertian reflectance of many surface covers which intimately couple the apparent reflectance to the environmental conditions under which the surface was observed.

As the atmospheric, solar and viewing conditions change both temporally and spatially, the magnitude of these radiometric effects changes dramatically. Therefore, in order to make any meaningful quantitative comparisons between data acquired at different times, at different places, or with different sensors, it is imperative that some form of radiometric normalization be performed.

3 RADIOMETRIC CORRECTION

Radiometric normalization must be performed to allow the meaningful comparison of multitemporal, multispatial or multisensor datasets. This normalization may be performed to some extent by statistical means; however such methods give no quantitative measure of the actual surface properties. Further, for applications such as change detection, it seems to defeat the purpose to statistically normalize two images so that one can tell what has changed.

Radiometric correction to surface reflectance on the other hand has the advantage of allowing quantitative comparison of the corrected data. Not only can one make relative statements about what has changed between two images, one can also state in absolute terms how much the reflectance has changed. Further, classification based on ground reflectance can be much more intelligent than a purely statistical classifier since the reflectance characteristics of the surface can be cross-referenced with a spectral database allowing not only surface classification, but surface identification.

To meet these needs, a prototype atmospheric correction system has been developed. The system takes as input radiometrically calibrated data and produces a reflectance map of the surface. The system is based on an analytic model of the atmosphere which is used to estimate the magnitude of the radiometric perturbations induced by environmental conditions. Having determined their magnitudes, the atmospheric model is inverted converting radiance at the satellite to reflectance of the surface.

The system is capable of operating in a automated environment. In this mode, all inputs are obtained using ancillary data together with the imagery itself. Alternatively, the user may interact with the correction process by providing additional environmental information. Variations in the atmospheric conditions over larger images are handled by introducing a spatially varying component into the atmo-

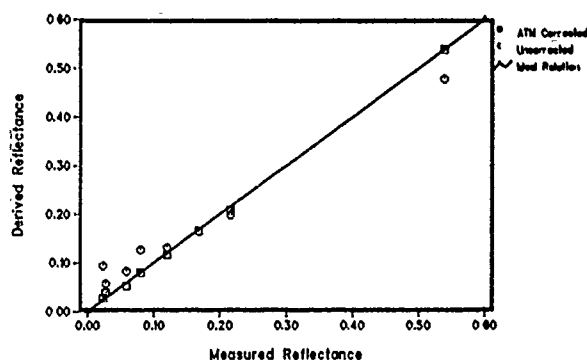


Figure 1: Satellite derived versus measured reflectance for uncorrected and atmospherically corrected Landsat TM data at the Maricopa Agricultural Center test sites.

spheric model. This allows for a more robust and accurate correction over full scenes. If a digital terrain model is available, it may be used to supplement the correction allowing for the compensation for effects introduced by local variations in the direct and diffuse solar irradiance of the surface.

4 ABSOLUTE ACCURACY ASSESSMENT

An important phase in the current work has been assessing the accuracy of the system under a variety of conditions to gain a quantitative feel for how well the correction is working. The results presented here include both absolute accuracy assessment, and relative accuracy assessments.

Perhaps the most important accuracy measurement of an atmospheric correction system is the absolute accuracy of the system—i.e. the ability of the system to retrieve surface reflectance in an operational environment. Unfortunately, this is also the most difficult measure to obtain as it requires having accurate surface observations at one or more test sites made close to the time of the satellite pass. With the help of Dr. P. N. Slater of the University of Arizona such a dataset was put together and used to estimate the end-to-end accuracy of the system.

The dataset itself consists of a radiometrically calibrated Landsat TM image of the Maricopa Agriculture Center in Arizona acquired July 23, 1985. A detailed set of simultaneous atmospheric measurements and ground reflectance measurements for two homogeneous fields (a cotton field and a bare soil field) was provided by Slater. The uncertainty in the reflectance measurements was well below one Landsat TM quantization unit.

The two test sites were first identified in the imagery using maps of the area, and the mean apparent reflectance at each site calculated in bands TM1 through TM4. Unfortunately surface measurements were not available for bands TM5 and TM7. Next, atmospheric effects were corrected for using the fully automated mode of operation, thus simulating conditions in a production environment. The uncorrected and corrected reflectance values were then plotted against the surface measurements made at each site and the results analyzed. The resulting plot is shown in Figure 1.

Analysis of this plot illustrates a number of interesting trends. As expected, the apparent reflectance of darker targets is higher than

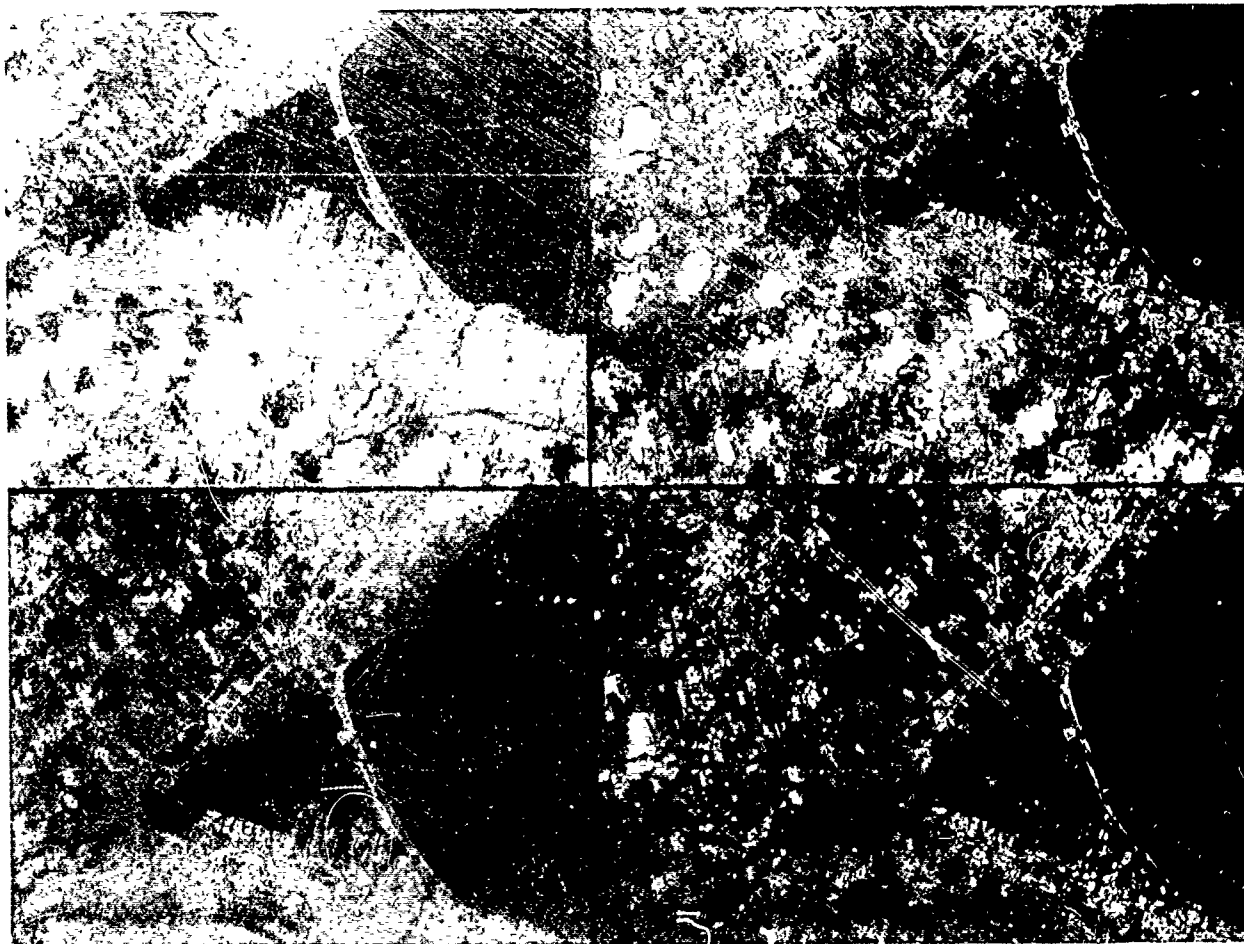


Figure 2: A black and white rendition of a false-colour composite image illustrates the difference between two uncorrected (left) and corrected (right) images of Hamilton, Ontario acquired August 3 (top) and September 20 1985 (bottom). Differences between the original images on the left clearly demonstrate the radiometric effects induced by the environmental conditions under which the imagery was acquired. After correcting for atmospherically induced effects, the two images on the right show greater contrast and correspondence, allowing easy identification of the changes between the two images.

their actual reflectance due mainly to the additive path radiance. High reflectance targets on the other hand appear darker in the uncorrected image due primarily to atmospheric attenuation of surface leaving radiation. The net result is a shift in the slope of the relation between derived versus actual reflectance from its ideal value of one. Correcting the data for environmental conditions compensates for these effects, resulting in better retrieval of surface reflectance. In all cases, the reflectance derived from the corrected image is closer to the actual surface reflectance, resulting in a more nearly ideal relationship between satellite-derived and measured reflectance values. In an RMS sense, atmospheric correction has reduced the total error from 0.039 reflectance units (RMS) in the uncorrected data to 0.01 reflectance units in the corrected data. This illustrates that, under reasonable atmospheric conditions, a production system running in an automated environment is capable of retrieving surface reflectance to within 0.01 reflectance units.

5 MULTITEMPORAL ACCURACY ASSESSMENT

For many applications, the temporal accuracy of an atmospheric correction system is of prime importance; that is, the question arises as to how well the system can normalize two multitemporal images. This is of particular interest in classification, where one is interested in extending spectral signatures between the datasets to, for example, eliminate retraining of the classifier.

To assess the multitemporal accuracy of the system, a dataset composed of a pair of Landsat TM images of Hamilton, Ontario and weather reports from the Canadian Atmospheric Environment Service was compiled. The two Hamilton images were acquired 48 days apart on August 3 and September 20 1985, and are representative of images with large variations in the state of the atmosphere. The August 3 image was acquired under clear atmospheric conditions with the hourly weather reports from Hamilton Station 'A' reporting a visibility of greater than 28 kilometers. The September 20 image, on the other hand, was acquired under hazy conditions with visibility reported at 13 kilometers.

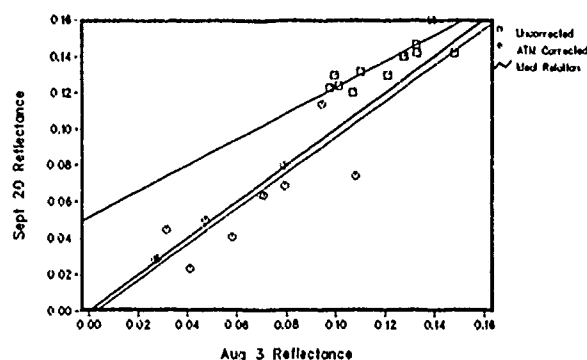


Figure 3: The reflectance of 12 test sites derived from an August 3 Landsat TM band 1 image of Hamilton is graphed against the reflectance of the same sites derived from a September 20 TM band 1 image. A linear regression through both the uncorrected values and the corrected values is shown along with a line representing the ideal relationship.

The images were independently corrected using the fully automated mode of operation. A black and white rendition of the uncorrected and corrected images is shown Figure 2. On the left side of this figure are the two uncorrected images. Notice the large difference between the two images, and in particular the poor contrast of the lower-left scene typical of images taken under the hazy conditions. The two corrected images shown on the right have much better contrast and similar dynamic ranges. The close radiometric matching between these two images allows for easier analysis and comparison of the similarities and changes in surface characteristics between the two dates.

To provide a more quantitative assessment of the improvement in radiometric registration, a total of 12 sites of unchanged reflectance were identified by viewing the images simultaneously on a video display. Care was taken to avoid areas such as agricultural fields which might have changed significantly over the 48-day period. Once identified, the apparent and corrected reflectance values were calculated for each test site. If the test sites truly were unchanged, a plot of the actual reflectance of the sites in August against their reflectance in September should show a strong linear relationship, with a slope of one and a zero intercept.

Figure 3 illustrates a plot of the August 3rd reflectance versus the September 20th reflectance for each of the 12 test sites in TM band 1. Examining this graph, it can be seen that the corrected reflectance closely approximates the conditions that would apply if the atmospheric effects had been compensated for exactly. By comparison, the uncorrected reflectance results differ significantly from the "ideal" relation illustrating the difference in atmospheric conditions between the two images.

Statistically, a regression line fit to the reflectance derived from the two dates should have a slope of one and intercept close zero if all atmospheric errors have been removed. Such regression was performed for the bands most effected by the atmosphere (TM bands 1, 2, and 3) the results of which are given in Table 1. The regression lines through the apparent reflectance data have a slope drastically different from one in all three bands clearly illustrating the effects of haze in the September image on the satellite observed signal. Regression lines through the corrected data however show significant improvement, having slopes and intercepts very close to the ideal values. Results reported in [Kauffman 89] show that similar improvements can be expected from multisensor datasets, thus allowing for the extension of spectral signatures between Landsat TM and SPOT HRV sensors. This suggests that the atmospheric correction procedures described

| Band | Correlation Coefficient | Uncorrected | | Corrected | |
|------|-------------------------|-------------|-----------|-----------|-----------|
| | | Slope | Intercept | Slope | Intercept |
| TM1 | 0.975 | 0.73 | 0.048 | 0.99 | -0.004 |
| TM2 | 0.988 | 0.81 | 0.026 | 1.06 | -0.004 |
| TM3 | 0.987 | 0.81 | 0.018 | 1.01 | -0.004 |

Table 1: Regression coefficients for a line fit to the reflectance of 12 unchanged test sites located in August 3 and September 20 images of Hamilton, Ontario. Lines fit through the corrected data are much closer to the ideal line having a slope of one and a zero intercept.

here will vastly improve the analysis of multitemporal multisensor datasets.

6 CONCLUSIONS

Non-systematic radiometric errors introduced by the environment must be compensated for before multitemporal or multisensor datasets can be compared. A prototype system has been developed capable of removing atmospherically induced perturbations from multispectral data by modelling the atmosphere and transforming radiance observed by the satellite into ground reflectance. Absolute accuracy assessment against ground reference data has shown the system is capable of retrieving surface reflectance to within 0.01 reflectance units (RMS). Relative assessment on multitemporal datasets has demonstrated a dramatic improvement in radiometric registration between images. The results of this analysis suggest that the system provides a viable method for radiometrically correcting multitemporal, multispatial, and multisensor datasets allowing the quantitative exploitation of remotely sensed data.

References

- [Ahern 87] F. J. Ahern, R. J. Brown, J. Cihlar, R. Gauthier, J. Murphy, R. A. Neville, and P. M. Teillet. Radiometric correction of visible and infrared remote sensing data at the Canada Centre for Remote Sensing. *Int. J. Remote Sensing*, 8(9):1349-1376, 1987.
- [Kauffman 89] D. Kauffman and B. Robertson. Extension of spectral signatures between multitemporal and multisensor datasets. In *Proceedings of IGARRS'89*, 1989.
- [Proy 86] C. Proy. *Integration du relief au traitement d'images de télédétection*. PhD thesis, L'Institut National Polytechnique de Toulouse, January 1986.
- [Slater 87] P. N. Slater, S. F. Biggar, R. G. Holm, R. D. Jackson, Y. Mao, M. S. Moran, J. M. Palmer, and B. Yaun. Reflectance and radiance-based methods for the in-flight absolute calibration of multispectral sensors. *Remote Sensing of Environment*, 22:11-37, 1987.
- [Tanré 81] D. Tanré, M. Herman, and P. Y. Deschamps. Influence of the background contribution upon space measurements of ground reflectance. *Applied Optics*, 20(20):3676-3684, October 1981.
- [Teillet 86] P. M. Teillet. Image correction for radiometric effects in remote sensing. *International Journal of Remote Sensing*, 7(12):1637-1651, December 1986.
- [Woodham 87] R. J. Woodham and M. H. Gray. An analytical method for radiometric correction of satellite multispectral scanner data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3):258-271, May 1987.

Atmospheric Correction of NIMBUS-7 CZCS Images Using Multiple Scattering Data Base

Sonoyo Mukai

Kanazawa Institute of Technology, Ishikawa 921, Japan

It is shown here that atmospheric correction of the NIMBUS-7 CZCS data based on the multiple scattering calculation in an atmosphere-rough surface model produces an precise and clear image of the current situations of the ocean.

In order to reduce the processing time of an image, the numerical results of the multiple scattering calculation are stored in the data base.

For more realistic atmospheric correction, the optical thickness of aerosol measured at National Institute for Environment Studies (in Tsukuba, Japan) is applied to process an image of the ocean near Japan.

Key words: Atmospheric correction, Multiple scattering, CZCS

1. Introduction

In the marine remote sensing, the intensity of radiation from the ocean is generally an order of magnitude less than the total intensity observed at the sensor equipped on the satellite. This is mainly caused from the multiple scattering of the solar radiation in the intervening atmosphere between the sensor and the sea surface, what one calls atmospheric effect. Removal of this atmospheric effect from the original remote sensing images is termed atmospheric correction (Gordon, 1978). This paper describes atmospheric correction of the Nimbus-7 CZCS data based on the multiple scattering calculations of the solar radiation in the atmosphere-rough surface system.

The Coastal Zone Color Scanner (CZCS) equipped on the satellite Nimbus-7 is a scanning radiometer which has six coregistered wavelength bands. Namely Bands 1, 2, 3, 4, 5 and 6 correspond to the wavelengths of 0.44, 0.52, 0.53, 0.67, 0.75 and 1.1 μm , respectively (Zion, 1983). The CZCS is aimed to estimate the phytoplankton pigment concentration near the sea surface by measuring the reflected radiation from the ocean.

A transfer problem is dealt with a superposition method in an atmosphere-rough surface model (Mukai, 1977). In the atmosphere, Rayleigh scattering by molecules and Mie scattering by aerosols are taken into account. The model of the Earth's atmosphere is constructed on the basis of Lowtran 6 code which gives the aerosol and molecular distribution with height (Kneizys et al. 1983). For a sea-surface model, the sea wave pattern given by Cox and Munk (1954) is adopted.

In order to reduce the processing time of an image, the numerical results of the multiple scattering calculation are stored in the data base named MSDB, which is short for Multiple Scattering Data Base.

For more realistic atmospheric correction, the optical thickness of aerosol measured at National Institute for Environment Studies (in Tsukuba, Japan) (Sasano, 1985) is applied to process an ocean image near Japan.

2. Atmospheric Correction

Fig. 1 represents a flow chart of our atmospheric correction, where subscripts a and g denote, respectively, aerosol and molecular gases. The numerical values of optical thickness τ_i and mixing ratio f_g are derived from LOWTRAN 6 code. The LOWTRAN 6 is a software program which calculate the atmospheric transmittance and radiance of single scattered sun light. Specifying the parameters of LOWTRAN 6 code, required values of τ_i and f_g are obtained.

On the other hand the aerosol optical thickness has been measured with a laser radar with wavelength $\lambda = 0.53 \mu\text{m}$, what one calls lidar, at National Institute for Environment Studies in Tsukuba, Japan (Sasano, 1985). We found that the average value of optical thickness of atmospheric aerosol above the Japanese island is less than that of standard middle latitude atmosphere derived from LOWTRAN 6.

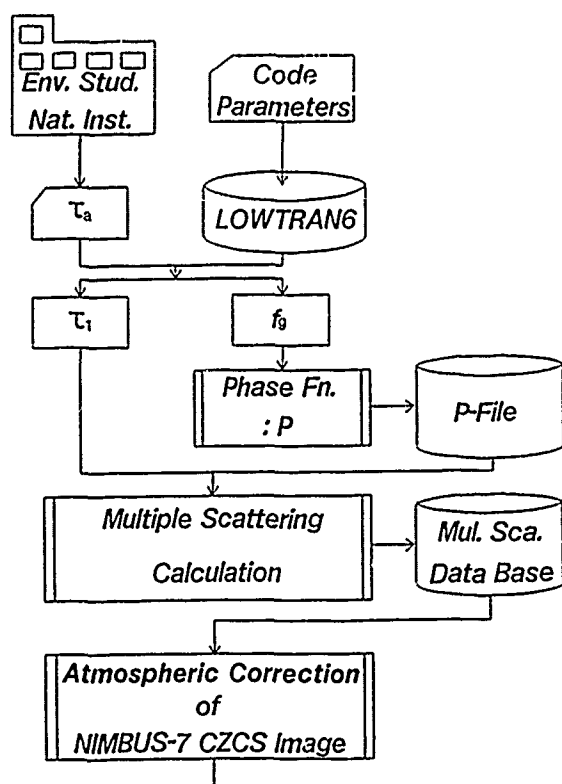


Fig.1: Flow chart of atmospheric correction.

A. Multiple Scattering Data Base (MSDB)

Putting function C on the ratio of the upward intensity just above the sea surface to that at the top of the atmosphere, which values are calculated from the equation of radiative transfer by using the superposition method. The ration C represents an effectiveness of radiation from the sea surface with respect to the total radiation reached to the sensor, or in other words a value of $(1-C)$ denotes the scattering efficiency of the atmosphere. We shall call this C-value by the coefficient of atmospheric correction. Because multiplying radiance of each pixel of original remote sensing image by this coefficient C produces an atmospheric corrected image. An algorithm of atmospheric correction proposed here is different from that defined by Gordon et al.(1987), however we can remove a certain kind of atmospheric effect.

Since the calculation of the multiple scattering in an atmosphere-sea surface model takes a long computing time, the numerical results of C are stored in the data base named MSDB (see Fig.2). Using this data base MSDB we can process a prompt atmospheric correction.

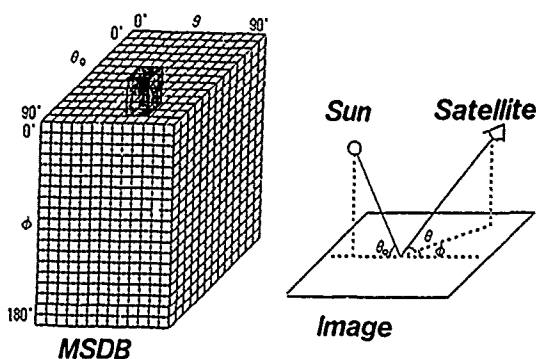


Fig.2: Diagram of Multiple Scattering Data Base.

B. Coastal Zone Color Scanner (CZCS) Images

Our results are applied to interpret the remote sensing data given by the Coastal Zone Color Scanner (CZCS) equipped on the satellite Nimbus-7, which has six wavelength bands of observation. An image in this paper is derived from CZCS Band 2 (wavelength of $0.52\mu\text{m}$) data observed on 14th April 1981 (orbit no.12477). Note that such bright areas as land and cloud are subtracted from an original image of Band 5 ($0.75\mu\text{m}$) and repainted in black, because we are now interested in the oceanic area.

Here an atmosphere-sea surface model is represented by four parameters, i.e. { 1.aerosol type, 2.optical thickness, 3.gas-aerosol mixing ration, 4.wind speed (m/sec)}. In terms of this notation, Fig.3 shows an image processed by atmospheric correction based on the model (Water-Soluble, $\tau_t=0.4$, $f_g=0.36$, $w=2$). The Water-Soluble aerosol model represents the absorbing particles with complex refractive index of $1.53-0.006i$.

Comparing with the processed image and the original one, it is found that atmospheric correction provides us with more precise information on the sea surface, e.g. intrinsic current situations. For the sake of comprehension, the radiance distribution along the white solid line in the image of Fig.3 is shown in Fig.4. Referring to Fig.5, which represents the presumed stream line between 1 April and 16 April(JODC, 1981). From these three figures 3, 4 and 5, our atmosphere-sea surface model can well express the current flow, at least concerned season and area.

3. Results

From the processed image of the Pacific Ocean near Japan, we obtain the following results;

- 1) The value of aerosol optical thickness measured in Japan produces a suitable image to the Black Current.
- 2) The absorbing particle called Water-Soluble is a good aerosol model above the Pacific Ocean.

In respect of an algorithm of atmospheric correction we draw the conclusions as below:

- 1) Multiple light scattering provides us with an accurate atmospheric correction.
- 2) Computational effort to process an image is reduced by using a data base stored calculated results.
- 3) In practice, reality of our algorithm depends on the model atmosphere-ocean.

References

1. Gordon, H.R., "Removal of Atmospheric Effects from Satellite Imagery of the Oceans," Appl. Opt. 17, 1631, 1978.
2. Zion, M., "Description of Algorithms for Processing Coastal Zone Color Scanner (CZCS) Data," JPL Publ. 83-98, NASA 1983.
3. Mukai, S., "Multiple Scattering in the Atmosphere with a Rough Surface," Astrophys. & Space Sci. 51, 165, 1977.
4. Kneizys, F.X., Shettle, E.P., Gallery, W.O., Chetwynd, J.H., Abreu, L.W. Jr., Selby, J.E.A., Clough, S.A. and Fenn, R.W. "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 6," Air Force Geophysics Laboratory, AFGL-TR-83-017 1983.
5. G Cox, C. and Munk, W., "Statistics of the Sea Surface Derived from Sun Glitter," J. Mar. Res. 13, 198, 1954.
6. Sasano, Y., "Observational Study on the Atmospheric Mixed Layer and Transition Layer Structure Using a Mie Lidar," J. Meteorol. Soc. Japan, 63, 419, 1985.
7. Gordon, H.R. and Castano, D.J. "The coastal Zone Color Scanner Atmospheric Correction Algorithm: Multiple Scattering Effects," Appl. Opt. 26, 2111, 1987.
8. Japan Oceanographic Data Center, "Ocean Currents Data," 56-8, 1981.

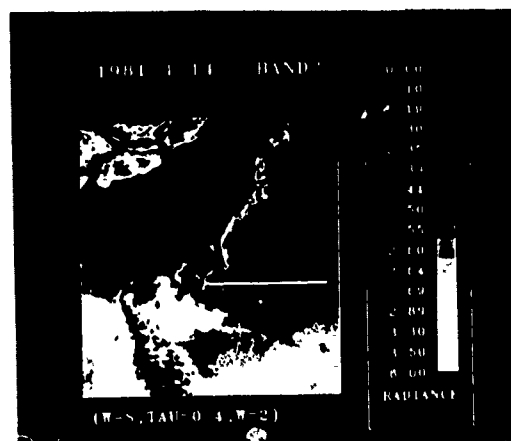


Fig.3: Image processed by atmospheric correction with the model (Water-Soluble, $\tau_1=0.4$, $f_s=0.36$, $w=2$)

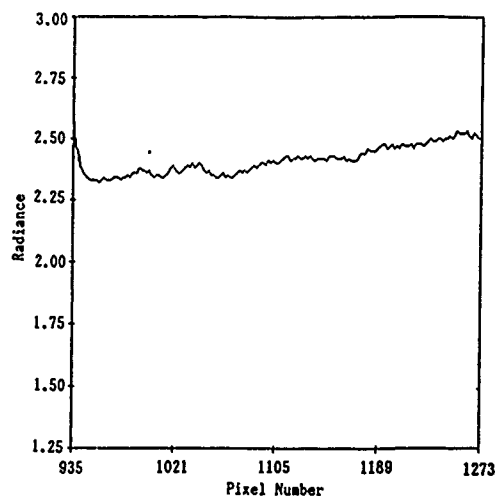


Fig.4: Radiance distribution along the solid lines in Figs.3.

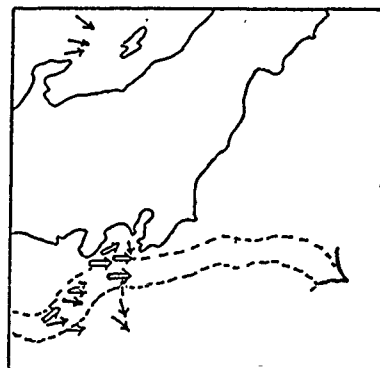


Fig.5: The presumed stream line between 1 April and 16 April.

USING LOWTRAN 6 AND DEM TO DERIVE PATH RADIANCE
FOR SPOT IMAGERIES OVER MOUNTAINEOUS TERRAIN

J.-Y. Chen, A. J. Chen and H. T. Wang
Center for Space and Remote Sensing Research
National Central University
Chung-Li, Taiwan, R. O. C.
Tel:886-3-4257232, Fax:886-3-4254908

A new approach is proposed to derive the altitude variation of path radiance for SPOT imageries over mountainous terrain. This represents one of the major steps to obtain a terrain reflectance distribution. As the water body has lowest reflectance for IR band, the IR radiance over the water body is commonly used for estimating the path radiance. However, in the case of the small lake surrounded by vegetation, the high reflectivity of the vegetation in near IR band causes the radiance above the lake to be brighter in the near IR than in the visible due to adjacency effect (Y. J. Kaufman, IEEE Trans. Geos. and Remote Sensing, 26, 441-450, 1988). On the other hand, the red band (0.61-0.68 μm) has very low reflectance for the vegetation and the water body and therefore, is better than near IR band for estimating path radiance.

The main feature of our approach is to get a better estimated value of meteorological range in LOWTRAN 6 code. The exterior orientation parameters of the SPOT satellite are, firstly, determined by space resection. By incorporating the determined orientation parameters with DEM and the satellite imagery, the ortho-image is, then, generated by photogrammetric techniques. The ortho-image is overlaid with DEM to find out the altitude profile of the minimum pixel values. Meanwhile, by using LOWTRAN 6 and the rawinsonde data taken at the same time, a set of altitude profiles of path radiance is obtained for different values of meteorological range. The best value of the meteorological range is determined when the associated path radiance profile intersects only once with the altitude profile of the minimum pixel values. Now that the path radiance is known at different altitudes, the path radiance associated with each pixel in the image is, thus, determined through the corresponding DEM. Two SPOT scenes taken on October 3 and 4, 1986 in northern Taiwan are analyzed for case study. The study area covers the size of 7.8 km x 10 km with the altitude range between 200 m and 770 m. The pixel values in small lake for IR band are either somewhat greater or smaller than those for red band. On the other hand, by using LOWTRAN 6, the path radiance of IR band is only about 50% that of red band. In other words, IR band can not be used for estimating path radiance. Our approach can resolve this problem. The minimum values of red band are found to be 18.38 $\text{w/m}^2 \text{sr-}\mu\text{m}$ at 350 meters and 33.84 $\text{w/m}^2 \text{sr-}\mu\text{m}$ at 320 meters for the first and second scenes, respectively and the corresponding meteorological ranges are 39 km and 12 km. The fact that the minimum pixel values are located at similar altitudes suggests the consistency of our approach. The path-radiance profile is also consistent with the corresponding rawinsonde data.

OPTICAL ARCHIVING FOR SCIENTIFIC DATA

Barbara E. Lowrey

Science Information Systems Center
 Space and Computing Division
 Goddard Space Flight Center Code 636
 Greenbelt, MD 20770

Abstract

Write-Once-Read-Many (WORM) optical disk technology is now practical for use as (1) a near on-line storage media, (2) long-term storage in a passive repository, and (3) under some circumstances, transporting from and to data-generators and data users at remote sites. Optical technology provides several significant advantages over magnetic tapes as a data storage media because (1) it is more compact, (2) provides random access, (3) has a longer lifetime, at least 30 years, and (4) is less sensitive to environmental factors such as temperature and humidity.

The Space and Computing Division now has 2 major archival projects using optical technology underway: (1) the Coastal Zone Color Scanner Data, and (2) the Dynamics Explorer Data. Both projects have used the Software for Optical Archive and Retrieval (SOAR) to transfer data in a volume format that can be read transparently by VAX/VMS. The software has been tested on optical drives by several manufacturers. The use of this software technique provides several advantages: (1) data generators and data users do not have to rewrite extent software and (2) the VAX/VMS Files-11 volume format is in wide use so that it serves as an interim "standard" and will be accessible in the long-term. A second software system, the Virtual Optical Disk (VOD), is under development; this provides a read/copy capability of files written in diverse formats such as UNIX, VAX/VMS, and the Space Telescope format.

Introduction

Four optical technologies have great potential for the storage of and ready access to image data. These are 1) WORM (Write Once Read Many) 2) CD-ROM (Compact Disk - Read Only Memory) 3) Erasable, or magneto-optic 4) optical paper. Of these, the first 2 are feasible and are coming into use for scientific data applications, magneto optic is eminent but there is yet a lack of field experience, and optical paper or tape is currently in prototype devices.

Experience has been gained during the last 4 years at Goddard with the use of WORM optical technology for data storage, and the use of CD-ROM has now begun. Both technologies are proving to increase the availability and utility of large science data sets. Remotely sensed images produce large data sets, and

for which both WORM and CD-ROM will be convenient, compact, and long-lived media. Because these media provide random access, operations such as image subsetting are facilitated.

The WORM and CD-ROM are complimentary technologies which have established growing market niches. Although the media consist of visually similar disks, there are fundamental differences in the physical creation of the disks of data and user actions taken to create the disks. Each technology has application-dependent advantages and disadvantages.

Applications

The Goddard applications of optical technologies began with the consideration of archiving and distributing data from the Dynamics Explorer (DE) satellite. DE is a 2-satellite system, launched in 1981, to make correlative measurements of solar-terrestrial interactions with a variety of experiments. The data bases associated with the diverse experiments are large, some of the size of 10s of Gigabytes, and have been prepared on DEC computers using VAX/VMS Files 11 structures. The University of Texas at Dallas proposed the use of WORM disks to archive the data and to facilitate correlative studies; UTD initiated development of a technique that would integrate the WORM drive into the VAX/VMS operating system. This resulted in the development of SOAR, Software for Optical Archive and Retrieval, by the National Space Science Data Center (NSSDC). At present, data bases at diverse DE sites are being transferred to optical disks for data interchange and archiving.

SOAR is a software system that provides the capability to produce a WORM disk in a magnetic disk emulation mode. This is enabled in several phases. The first phase is the SOAR initialization of the optical disk. This establishes a "pseudo-device" with the directory to the files located on magnetic disk and reserves an area on the optical disk for the eventual location of the directory. The second phase is the write cycle of the optical disk. As files are copied to build the optical disk, the directory is updated on magnetic disk to understand the new files; this updating requires overwriting that can not be done on the write-once media. In this phase, SOAR is needed to perform the write operations; however, many ordinary DCL commands addressing the pseudo-device are used, including DCL Read, directory inquiries and DCL Copy

from optical to magnetic disk. The final operation is to "close" the optical disk; this transfers the directory on magnetic file to the space reserved on the disk and disestablishes the pseudo-device. Thereafter the disk and drive are operated as any other VAX/VMS Read-only peripheral; the full directory, subdirectory and file structure of Files 11 is available to the user.

The present version of SOAR contains some valuable enhancements. A SOAR label and a label checking option is available for the optical disk to prevent a write operations going on to the wrong disk. The next version of SOAR will include the ability to store directories "temporarily" on the optical disk in order to serve as a backup in the event of corruption of the magnetic disk and to facilitate interchange of unclosed disks between sites.

The Coastal Zonal Color Scanner (CZCS) was launched on Nimbus-7 in order to determine the feasibility of measuring biological characteristics of coastal and ocean basin waters. This data proved useful for a variety of scientific questions and advances in data processing technology provided capability, so NASA decided to process and archive all useful data taken during the eight years of instrument operation. About 700 gigabytes of CZCS data on 28,000 tapes have been transferred to 220 optical disks for improved accessibility and to preserve the data. From these data it was then feasible to assemble a composite image showing the global biological activity in the oceans at a resolution of 20 km, compiled from the original 1 km pixels. (Feldman et al, EOS, in press). This was accomplished by means of an optical jukebox, using SOAR to provide the interface to the host VAX and some additional software to operate the jukebox. Creation of this image would not have been possible when Nimbus 7 was launched in 1978.

In the CD-ROM area, Jet Propulsion Laboratory (JPL) has already distributed widely its first CD-ROM of planetary image data. The NSSDC has acquired a CD ROM premastering system and has produced its first CD-ROM, an astronomical data base that had hitherto been stored on a mainframe. A test disk of Comet Halley data is being prepared. CD-ROM is planned as the distribution mode to the international Halley experimenters.

1. Operational Scenarios for CD-ROM and WORM

Although the media (disks) look very similar for both CD-ROM and WORM, the processes to generate data containing disks are fundamentally different. A good analogy is that WORM compares with CD-ROM as typing a manuscript compares with printing a book. The WORM is operated by a science user writing via a host computer in much the same manner as a science user writes to a tape drive or a magnetic disk generating one WORM disk at a time. For CD-ROM, data is prepared at the host computer site, and shipped on magnetic tapes to an external facility. The external facility replicates CD-ROMs in quantity.

Data is written on a WORM disk by a WORM drive attached to the science data user's host computer. The laser head of the WORM drive is narrowly focused to heat a small spot on the media and alter that spot permanently. The equipment is operated much the same as any other computer peripheral. The salient difference between WORM and magnetic drives is that some type of additional software integration into the host operating system is necessary because of the Write-Once nature of the media, specifically, because

directory information cannot be altered as files are added; often, either tape emulation or magnetic emulation is used. Tape emulation mode facilitates interchange between diverse host computers.

CD-ROM disks are prepared in a series of steps. First, data to be put on CD-ROMs is prepared in a standard volume format, ISO 9660, by (commercial) premastering software. This step may be done by an external supplier who has received digital data tapes from the science data generator. Then, the volume formatted data is shipped on magnetic tapes to a mastering facility. The facility generates a master and produces many CD-ROMs from the master.

Worm Technology

The sizes of WORM drives were standardized to the footprint of a floppy disk and drives. The major manufacturers have been active mainly with 5 1/4" and 12" drives and media, with some interest in 14".

As of this writing, there has been no further progress toward formal standards for 12" drives. The lack of standards means that even blank disks must be ordered specifically for a drive (unlike floppy disks), although market pressures are acting to reduce divergence somewhat. Additional standards appear to be eminent for 5 1/4" drives.

The amount of data that can be put on a disk is directly proportional to the size of the disk, because the ability to alter a surface depends on the power of the laser diode that writes to the disk. The current generation of optical disks permit 200 Megabytes on a 5 1/4" disk, and 1 Gigabyte on a 12" disk, although a Sony can hold 1.6 Gigabytes when operated in the CLV mode (Constant Linear Velocity). The 14" disk holds 3.4 Megabytes/side.

CD-ROM

The Compact Disk - Read Only Memory derives from the success of the CD-Audio market for the distribution of recorded music. The development of a standard by major vendors was and is influential in the development of the media as a publishing media. A second major driver for the development of the standards has been the software industry. Thus, the CD-ROM now has both physical and logical standards that are unprecedented in a new technology. A CD-ROM is a 5 1/4" disk containing 660 Megabytes and is read by a CD-ROM player. The volume format is specified in ISO 9660: 1988 (E), "Information processing -- Volume and file structure of CD-ROM for information interchange"

To date, almost all of the applications and interfaces to read data files have been written on IBM PC class computers. With the announced support of both DEC and Apple Computer for ISO 9660, this may change soon.

CD-ROM disks are read by CD-ROM players. These are Read-Only devices and are very similar to CD Audio players used to play recorded music. In fact, the mechanical components and error correction codes of these two types of players are common. Currently, CD-ROM players cost from \$600 to \$1000; the "bottom" price of \$249/drive is projected soon. If a SCSI port is not available on the host computer, an personal computer class SCSI controller costs an additional \$600 and a VAX class about \$2000.

Mastering and replication are available from several facilities; mastering a CD-ROM currently costs about

\$1400 ; each disks replicated costs about \$2 (for a turnaround time of 5 days). These costs have dropped considerably in the last year due to the benefits resulting from a larger market. The overall cost per disk decreases with increased quantity, thus CD-ROM is a publishing media, with the overhead costs of mastering (and user data preparation) offset by quantity production of disks.

Future

The next generation of WORM drives and disks is eminent. These will have "double-density, double-speed" capacities. The performance and cost parameters for 12" such as, write data rate, read data rate, seek access time, and capacities, will vary according to manufacturer, so that selection may be application dependent. Probably the 12" size will continue to be the most significant size for science applications.

The production of CD-ROMs is expected to continue tripling each year. This will provide an extremely convenient and inexpensive means to mass distribute scientific data, including to university and general public libraries. Each Megabyte so distributed costs less than a cent!

Magneto-optic technology will soon be widely available and will fill a market niche, particularly in the 5 1/4". The erasable ability of this media will ease system integration problems.

"Multifunction" 5 1/4" drives are under development. These will have the capability of reading CD-ROM, magneto-optic, and WORM media, and writing magneto-optic and WORM media.

The "optical paper" drive may be significant for the capture of data from the new generation of satellites, as the plans are to produce drives that hold one tape containing 1 Terabyte of data, with an average access time of 22 secs and a sustained transfer rate of 24 Megabits/sec.

The powerful capabilities of these optical technologies can expedite image storage and retrieval remarkably

OPTICAL STORAGE SYSTEMS AS THE MEDIA OF REMOTELY SENSED DATA EXCHANGE

Joji Iisaka

Canada Centre for Remote Sensing
Energy, Mines and Resources Canada
Ottawa, Ontario, CANADA, K1A 0Y7

ABSTRACT

Computer compatible tapes are widely used as the media to exchange remotely sensed data. Due to the spatial and spectral resolution enhancement of remote sensors, the amount of data volume per scene has very much increased. As well, microcomputer-based image analysis systems have become very popular during the last several years, but the CCT is not a convenient media for these users of remote sensing.

This paper describes some experience in the use of an optical disk storage system such as CD-ROM and the 5¼" WORM (Write-Once-Read-Many) to exchange remotely sensed data at CCRS.

1. INTRODUCTION

Computer compatible tapes are commonly used by central institutions which receive the satellite data as the media to distribute remotely sensed data to users. The CCT is still the only media which can store large amounts of data and which has an industry standard compatibility between various tape drives. Due to the enhancement of spatial and spectral resolution, the total data of a scene has exceeded the current capacity of magnetic tapes, and several reels of magnetic tapes are now required to cover a typical scene with, for instance, LANDSAT TM (Table 1).

Remote sensing data on CCTs is used mostly at the operational institutes of remote sensing which would have relatively larger computer facilities with many magnetic tape drives. Recently, many microcomputer-based image analysis systems have become popular, but CCTs are not a convenient media to these users.

Recent technology advancement in optical disk storage systems such as CD-ROM and the 5¼" WORM, which have large storage capacities such as 300-800MB, could resolve these media problems for remote sensing. Drive costs are also affordable to both microcomputer-based and main/mini computer-based users.

At CCRS, we have examined this technology as the data distribution media for remotely sensed data.

This paper will discuss both our experiences in this technology, and the proposal for a data format for remotely sensed data with DASD (Direct Access Storage Device).

2. SMALL OPTICAL DISK STORAGE SYSTEM AT LDIAS

At present, there is a wide variety of optical disk storage technology available: CD-ROM, 12" WORM, 8" WORM, and 5¼" WORM are examples of available media today. Some type of optical disks are already being used for image data bases, because of their huge storage capacity.

For the purpose of remotely sensed data and information exchange, compatibility of media and drives, as well as cost of data replication and distribution, are key factors to the users of remote sensing independent of whether they are using mainframe computers or microcomputer-based systems. From these points, we have investigated small optical disk technology and partly implemented it on LDIAS (LANDSAT Digital Image Analysis System) at CCRS (Goodenough *et al.*, 1988).

2.1 CD-ROM Experience

CD-ROM drives were marketed first as small optical disk systems which at least had physical standardization to some extent. Therefore, we conducted an experiment to use a CD-ROM as a medium for remotely sensed data exchange (Iisaka, 1987).

The advantages of CD-ROM were: (1) Large file capacity. Typically, a CD-ROM can store 550MB and this was large enough to store a full TM scene. (2) Low drive cost. The price range was affordable to the users of microcomputer-based image analysis. (3) Standardization of CD-ROM technology. When we started experiments on CD-ROM, there was only physical standards, but logical standards, i.e. standardization of volume and file format was later established as ISO# 9660.3, (ISO Manual, 1988). Some popular operating systems for microcomputers provided file management functions to read CD-ROM (Microsoft, 1986). (4) Reliability of media. Conventional CCTs were sensitive to mechanical shock and electric fading phenomena. (5) Space and mailing cost saving. Due to the compact size of a CD-ROM, storage space and mailing costs could be dramatically reduced.

As already mentioned, when we conducted our experiment with the CD-ROM, there was no standard logical format. Our sample CD-ROM was formatted in conventional MS-DOS format (Version 3.x) after dividing whole CD-ROM space into several volumes having logically the same format as the conventional file format of the day.

Due to CLV (Constant Linear Velocity) type rotation mechanism of CD-ROM drive, data transfer rate was slower than conventional hard disks (Table 2), but our experiment took about 30 seconds to display a 512 × 512 of 3 channels from our CD-ROM. This display time was quite acceptable compared with CCT, which required loading, record skipping, rewinding and volume handling times.

The major disadvantage of CD-ROM was the cost of initial mastering. CD-ROM replication processes assumes mass replication demands. Unfortunately, replication demands of remote sensing data for a specific scene is not too great so as to take advantage of this low replication cost.

Fortunately, writable CD has emerged which has write-once capability and can be read by conventional CD-ROM drives. This technology might resolve this disadvantage of the CD-ROM and also take full advantage of standardized CD functions (Hamada *et al.*, 1989).

2.2 5¼" WORM Disks

After completing investigations of CD-ROM technology, WORM (Write-Once-Read-Many) technology was examined. Currently, the WORM medium is available with 5¼", 8", 12", or 14" disks and smaller disks are under development.

Unfortunately, when we started this investigation, very little compatibility had been established for this medium, except for the physical specification of the 5¼" WORM cartridge. This was one of the reasons why we focused only on the 5¼" WORM. Physical standards for WORM drives have now been established for the 5¼" media.

In CCRS, 5¼" WORM technology was investigated for two purposes: one was the WORM media as private disk storage and the second was the WORM as the media for data exchange between main/mini frame computer and microcomputer-based systems.

Ideally, the WORM cartridge which is now used routinely as a private disk on LDIAS should be readable by other systems such as microcomputer-based systems. Much of the logical format information about conventional disks are proprietary to manufacturers. There is also no format standard of remote sensing data for DASD. So we determined to use existing data format for remote sensing, i.e., LANDSAT LGSOWG format.

Remote sensing data on CCTs were read onto the WORM cartridge through MAS-100, which emulates a tape drive as tape images on an as-it-is basis. We are now able to download and upload data sets from the LDIAS (VAX-11/780) and a full TM scene could now be stored on a side of WORM cartridge. The WORM cartridge storing the CCT images is transferred by hand to a microcomputer-based image analysis system which has its own drive. Image extraction programs are now employed only for further memory bound analysis, to the limits of the workspace available.

2.3 Results of Experiments

Data downloading time to a WORM through the MAS-100 was about 11 minutes per a CCT (1600 bpi): 5 minutes for tape loading, rewinding and unloading and about 6 minutes for data copy.

Sub-image extraction time at the IBM PC/AT from a WORM drive took 1 sec./line for 3 channels: therefore, to extract 512 lines, it took 9 minutes.

Through our experiment, we were able to conclude that WORM media were much more convenient than conventional CCTs and with their portability, an ideal media for input into image analysis systems for remote sensing. Table 2 contains WORM and CD-ROM specifications.

3. DATA FORMAT CONSIDERATION FOR OPTICAL DISK STORAGE

In our experiment, no advantages in data processing using DASD functions were considered, such as random access. Basically, the 5¼" WORM is a Direct Access Storage Device (DASD). With the advantages of standards for the physical media and the logical format for volume and file structure, these have been established through the efforts of the media and information processing industries. The task which remains in the remote sensing community is to standardize application-oriented data format of remote sensing data processing.

Before discussing a data format for remote sensing, a summary of the fundamental requirements for a data format applicable to the exchange of remote sensing data will be given: (1) All necessary information should be stored. Types of information required depends, however, on the user's level and application. Some users require detail information about sensor calibration, but some only need image data with minimum header information. (2) The format must be capable of storing large amounts of image data, and it must be flexible to data volume changes. Most existing image analysis systems require definition of storage space allocation in advance and this is an inconvenience. (3) Compatibility is another important issue regarding data format. Data format information should not be proprietary and should be published with easy access by the public. Ideally, the media for the distribution of remote sensing data should be compatible with all the media and reading devices manufactured by the different companies. (4) The logical format should be acceptable and upward compatible with popular operating systems. (5) The format should be flexible allowing future expansion causing minimum impact to existing systems. Addition of new types of information after requires a considerable effort to modify existing image analysis software. The need for program modifications and conversion should be minimized. (6) The format must be easy for data export and import. Most existing image analysis systems for remote sensing use their own internal image data format and provide functions to feed remote sensing data from CCT, but few data export and import functions are provided between image analysis systems and other data processing systems such as GIS and statistical analysis systems.

Conventional CCT format for remote sensing such as LGSOWG TM data format the volume content and files were specified by remote sensing users (Murphy, 1982), and consequently, most of these file management functions are now provided by every applications operating system today.

Locations of files, sizes of files, date of file creation and types of files are managed by conventional operating system functions, but data access frequency is dependent on the type of information. For example, image data itself will be accessed most frequently during the course of image analysis, and image header information is needed only at the beginning of image analysis sessions. Image look-up tables are required just before image display. It is therefore not necessary to include all information of different types in an image file thus increasing transaction speeds.

Proposed here is a data format of remote sensing applicable to optical disks as DASDs, and the following approach has been taken:

1. Fully employ the available file management functions provided by popular operating systems.
2. Separate an image data set into different files according to data type and characteristics. Header information which stores image and pixel sizes in a pixel could be stored separately from image data.
3. Use hierarchical directory functions to group specific image data sets: a directory at higher levels might be defined according to the sensor type employed such as LANDSAT TM data, MSS data and SPOT HRV. This directory might alternatively be defined according to a specific geographic location where the data set consisting of multi-sensor (pictorial) and data like maps (graphical) are grouped. This level of directory could also represent a specific project name. The next level of directory could be defined by, for example, a specific scene number, sub-image numbers, or image control name and identification. Of course, the root directory or the volume name could be used for scene identification, if the media does not store many different scenes.

By employing this hierarchical directory structure, specific data of interest can be easily grouped.

4. Standardization applied just to naming conventions for files and file extension. Some naming conventions are suggested in Table 3.

5. By employing the simplest data structure and data format, a digital image of two dimension would be presented by a two-dimensional matrix of rows and columns. Most of the image analysis programs therefore would only process this data type.

Information about matrix or the digital image size can be stored in a separated file as an image index file or a header file. A sub-image could be stored also in matrix form at the same level as the parent image. But the sub-image would have a different image index file which stores sub-image size as well as the location of the sub-image (Figure 2). If users know image sizes or image size is fixed, user needs only this image body file.

In this way, modification of header contents would give minimum impact to the existing image analysis program modules. The basic image file should store only the single band image. Multi-band image files such as BIL and BSQ format may be created with specific file extension names, but these files should be interpreted not as basic image files but as a file which was combined a set of basic files.

A file extension can be used to identify which spectral band is stored in that file such as .BN4, and a file storing information about image characteristics such as image size, bits size of a pixel and spectral band ranges might have file extensions like .IND.

This image index concept has been implemented on MCBIAS, and as an image index, it is not a large file, and information content is specific and simple, thus the file could be created using conventional editor functions of the operating system.

In Table 3, the proposed image data file concept is shown. Advantages of this format are: 1) flexibility to image size and the addition of new information; 2) ease to grouping and regrouping data; 3) minimal impact on the existing image analysis package; 4) ease in identifying required information for a volume; 5) ease of export and transport of data to other systems.

Complete details of data formats for each record have not been defined yet. The concepts to separate image data from other data information to give data portability between different image analysis systems has been partially implemented.

4. CONCLUSION

The optical disk storage system is definitely a promising technology for information and data exchange of remote sensing, and which might substitute for conventional CCT media. Data format compatibility needs to be established taking into account file management functions provided by conventional operating systems of host computers. The suggestion of a good starting point to standardize record format of each file with structured data format concepts is proposed here.

5. ACKNOWLEDGEMENT

The author wishes to thank Mitsui Petro Chemical Industry Co. for their assistance to investigate WORM technology.

6. REFERENCES

- (1) Iisaka, J., "Optical Disk Technology and Microcomputer-Based Image Analysis Systems", Proc. 40th Annual Conference of SPIE, pp. 199-204 (1987), Rochester, U.S.A.
- (2) "Sample Remote Sensing Images on CD-ROM User's Guide", MCBIAS.TM.87-001.02, Technical Memo, CCRS (1987), Ottawa, Canada.
- (3) International Standard ISO9660, "Information Processing-Volume and File Structure of CD-ROM for Information Interchange", ISO9660:1988(E), International Organization for Standard, Switzerland.
- (4) "Installation Guide for MS-DOS Extensions", Microsoft Corporation Document: 000080004-100-000-1186 (1986), U.S.A.
- (5) "ESA/EPO Optical Memory Report", European Space Agency Report ER-RP-EPO-GA-0202 (1988), ESA, Esrin, Italy.
- (6) Murphy, J.M., "Standard LANDSAT TM CCT Format", Technical Memorandum No. DMD-TM-82-249A (1982), CCRS, Canada.
- (7) Hamada, E., S. Yukia, and I. Takeshi, "That's CD-R", 1989 Technical Digest Series, Vol. 1, pp. 45-48 (1989) Optical Society of America, Washington, D.C., U.S.A. (Erasable MO disk)
- (8) Goodenough, D.G. and A. Ménard, "User's Guide to LDIAS, CCRS, Ottawa, 1988.
- (9) "Optical Archiving System MAS-100", Reference manual. Aquidneck System International (1988), Rhode Island, U.S.A.

TABLE 1: Data Volume of Typical Remotely Sensed Data

| | MSS | TTM | HRV |
|------------------------------------|------------------|------------------|------------------|
| Full Scene | | | |
| -Storage required by a single band | 8MB | 35MB | 9MB + 36MB |
| -Number of bands | 4 | 7 | 3 + 1 |
| -Storage required by all bands | 32MB | 245MB | 63MB |
| -Minimum number of 1600 bpi CCTs | 1 | 10 | 2 |
| 512 x 512 Sub-scene | | | |
| -Storage required by a single band | 0.25MB | 0.25MB | 0.25MB |
| -Storage required by all bands | 1.00MB (4 bands) | 1.50MB (6 bands) | 0.45MB (3 bands) |

TABLE 2: CD-ROM and 5¼" WORM Characteristics

| | CD-ROM | 5¼" WORM |
|--------------------|--------------------|--------------------|
| Memory capacity | 552 MB | 800 MB |
| Data block size | 2048 bytes | 1024 bytes |
| Access time | 1 sec. | 0.12 sec. |
| Rotation control | CLV | CAV/MCAV |
| Data transfer rate | 176 KB/sec. | 1.25 MB/sec. |
| Data error rate | <10 ⁻¹² | <10 ⁻¹² |

TABLE 3: Hierarchical Data Format of Optical Disks for Remote Sensing and Naming Convention

Directory

Sub-Directory

Examples of Name

1. Sensor type
2. Geographical Location
3. Project
4. Scene ID

File Name

1. Scene ID
2. Project
3. Geographical Location
4. Title of File Content (ex: a specific type of Header Format)
5. User Convention

File Name Extension

1. Image Data and Type
(Ex: .BN4 band 4 single image
.BN7 band 7 single image
.HB4 band 4 image in 16 bits
.BIL BIL data format with Index file of file extension .IND
.BSQ BSQ image data with an index file of file extension .IND)
2. Image Index File
(ex: .IND this file includes image size and Bits per bite, window area and members of band images)
3. Header File
(ex: .HDR File name may specify specific header type)
LGSWG.HDR stands for master header information of LGSWG data format
4. Look-Up Table
(ex: .LUT Look-Up table for that file
.LTR Look-Up table for red gun
.LTB Look-Up table for blue gun
.LTG Look-Up table for green gun)
5. Calibration Parameter File - .CAL
6. Image Pattern Mask File
(ex: .MCK training fields, etc.)
7. Histogram of image files with extension .HIS
8. Spectral statistics files with extension .STA
9. Any type of files can be added. For example, image description files with .TXT.

FIGURE 1. Optical Disks for LDIAS and MCBIAS

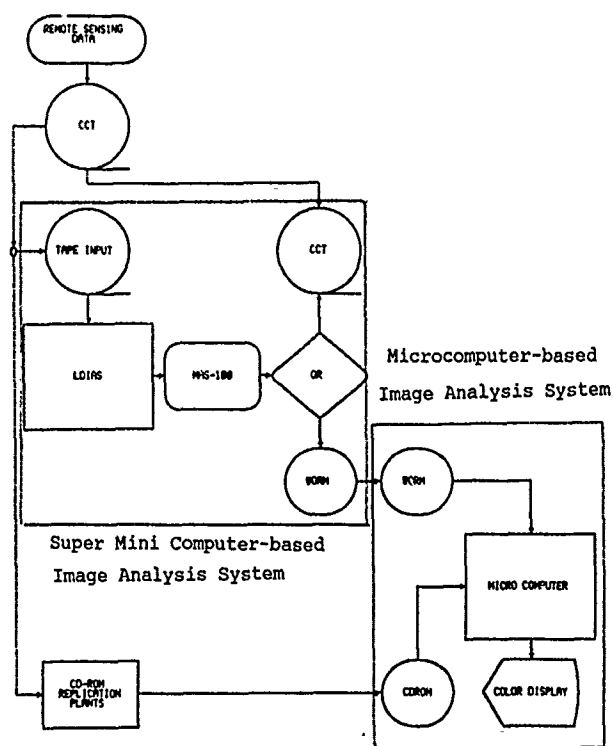
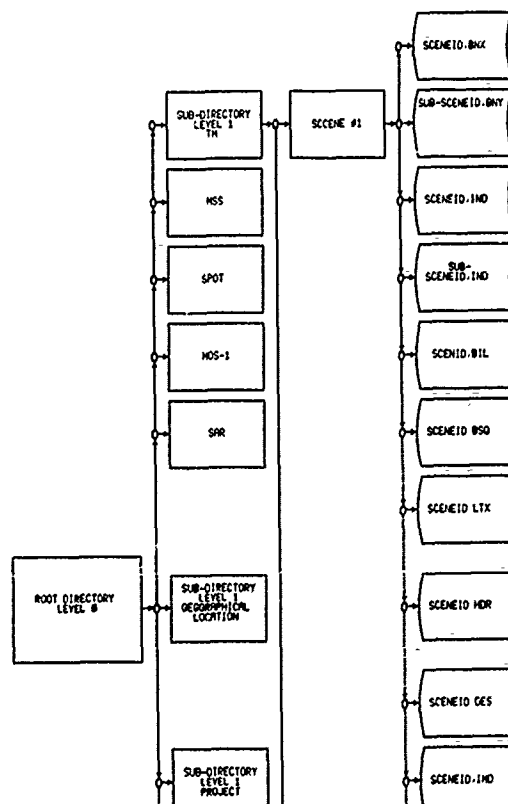


FIGURE 2: Hierarchical Image Data Structure



LANDSAT ELECTRON BEAM RECORDER
 IGARSS '89
 12th Canadian Symposium on Remote Sensing

P.F. Grosso, J.P. Whitley

Image Graphics, Inc.
 917 Bridgeport Avenue
 Shelton, CT 06484
 (203) 926-0100
 (203) 926-9705 - FAX

INTRODUCTION

Improvements in electron beam film recording technology make possible high quality, high resolution satellite and aerial imagery at throughput rates which are an order of magnitude faster than other recording techniques. The flexibility of an electron beam recorder when coupled with a mini or microcomputer controller provides an extremely cost effective recording solution for geometrically and radiometrically corrected satellite and aerial imagery for receiving stations or image processing workstations.

LANDSAT AND SPOT IMAGE GENERATION

A minicomputer controlled Electron Beam Recorder (EBR) is presently in use at the Brazilian Government's Instituto De Pesquisas Espaciais (INPE) satellite ground station. This 5 inch film size EBR was provided by Image Graphics, Inc. and is used to record both LANDSAT and SPOT satellite imagery in South America. Table 1 gives the performance characteristics of the Brazil LANDSAT SPOT EBR.

TABLE 1

BRAZIL LANDSAT/SPOT EBR PERFORMANCE

| | |
|--------------------------------|---------------------------|
| FILM SIZE | 127MM (5") WIDE |
| IMAGE FORMAT | 115 X 136MM (4.5" X 5.4") |
| ADDRESSABILITY | 65K X 65K |
| RESOLUTION | 18,000 X 21,000 PIXELS |
| BEAM DIAMETER | 5UM (0.0002") |
| VIDEO BANDWIDTH | 10MHZ |
| RASTER RECORDING RATE | 2.5 MEGAPIXELS/SEC |
| DENSITY RANGE DENSITY | 0.1 TO 2.1 DU |
| DENSITY UNIFORMITY | .02 DU |
| DEFLECTION JITTER | ±2UM |
| SCAN LINE SPACING UNIFORMITY | 1UM RMS |
| GEOMETRIC FIDELITY | 0.01% |
| CONGRUITY OF SEQUENTIAL IMAGES | SUM (0.003%) |

Each spectral band from the LANDSAT or SPOT image sensors are recorded as latent images on 5 inch wide silver halide film in the EBR and then processed in a conventional manner. Correction coefficients for geometric errors caused by satellite orbit, earth's rotation and sensor configuration and radiometric calibration for sensor errors and recording film are introduced during the recording. The range and degree of computer control corrections available while recording are listed in Table 2.

TABLE 2
BRAZIL EBR COMPUTER CONTROLLED CORRECTIONS DURING RECORDING

| CORRECTION OR CONTROL | RANGE | DISCRETE LEVELS |
|--------------------------|---------------------------------------|-----------------|
| EXPOSURE (FINE) | $\pm 10\%$ | 64 |
| VERTICAL OFFSET (COARSE) | $\pm 10.24\text{MM}$ | 4096 |
| VERTICAL OFFSET (FINE) | $\pm 1.28\text{MM}$ | 4096 |
| HORIZONTAL OFFSET | $\pm 10.24\text{MM}$ | 4096 |
| SKEW | $\pm 10\%$ ($\pm \text{TAN } -1.1$) | 4096 |
| LINE WIDTH (SPOT WOBBLE) | 16UM TO 18UM | 256 |
| VIDEO POLARITY | POSITIVE OR NEGATIVE | 2 |
| VIDEO CONTROL | ----- | 256 |
| VIDEO TRANSFER | LINEAR OR GAMMA | 16 |
| | CORRECTED (.5 TO 2) | |

Continuous tone composite color images are produced by exposing selected spectral bands of imagery through a color filter onto color film or paper. Each spectral band is punched in the EBR during recording to provide pin registration. The high resolution, recording accuracy and repeatability of the EBR enables the pin registered spectral bands to be optically enlarged 10 times using color filters to expose 40 inch color film and retain the fine spatial resolution of the original.

The recording flexibility of the EBR enables a significant ground station cost savings. The on line corrections introduced during the recording enabled corrected output recordings, using a mini-computer controller instead of requiring a main frame, for preprocessing and merging the image data to be recorded. A schematic overview of the EBR system is shown in Figure 1.

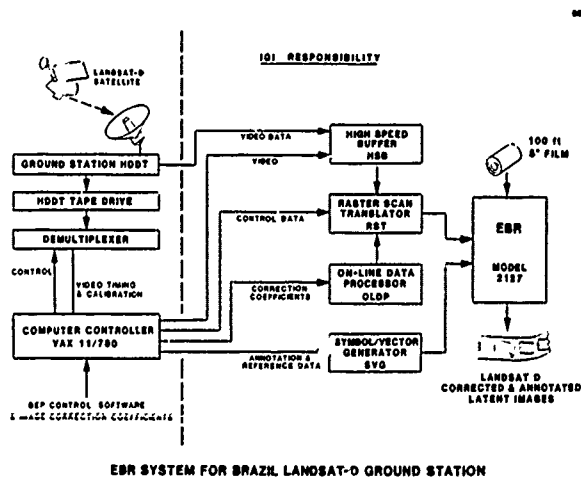


Figure 1. EBR FUNCTIONAL AREAS

IMAGE PROCESSING WORKSTATION RECORDING

The EBR has additional capabilities which enable the recording of data processed by image processing workstations. The EBR gamma or image density recording range may be expanded or compressed over any portion to match that used by an image processor to record the entire view of the overall LANDSAT or SPOT IMAGE as color separation or spectral band imagery. Because of the very high resolution, a 5 x 5 block (or more) of LANDSAT or SPOT IMAGES which have been digitally merged can be recorded for a synoptic area.

EBR TECHNOLOGY REVIEW

The electron beam recorder is analogous to a faceless cathode ray tube recorder where the film transport has been placed inside the vacuum. Exposing the film directly with electrons, which eliminates the phosphor-coated faceplate of the CRT and the lens which projects the image on the faceplate of the CRT to film, removed the major causes of loss of image quality (resolution, and edge acuity) and low exposure intensity which effects recording speed. The direct exposure of silver halide film from a high energy (15-20kv) electron beam has thousands of times more efficiency than exposure by light or laser beam. This enables extremely high data recording rates for cost effective throughput.

Figure 2 is a schematic layout of a typical EBR.

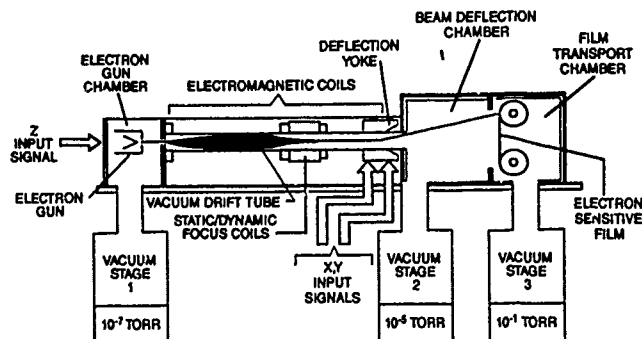


Figure 2. SCHEMATIC LAYOUT OF TYPICAL EBR

The electron gun assembly is a triode structure with a directly heated thermionic tungsten cathode. Small apertures in the grid and anode form an electron beam with a low divergence angle. The electron beam is accelerated at 15-20 Kilovolts through the vacuum drift tube and through the center of the electromagnetic focus and deflection coils. The beam is focused by the static focus coil into the small spot, typically 4-5 microns diameter at the film plane which can be used to produce an image on a 5" x 5" format in excess of 30,000 x 30,000 pixels. Dynamic focus, dynamic astigmatism and geometric correction circuits are used to correct for spot shape and image geometry as the beam is deflected with the deflection yoke over the biaxis image format.

The electron gun, the film and the film transport are housed in a three stage, differentially pumped, fully automatic, fail safe, vacuum system. The three stage vacuum system supports an excellent vacuum for the electron gun chamber which enables thousands of hours of operating life by the tungsten cathode; a good vacuum at the film plane thereby avoiding scattering of the electron beam by gas molecules; and a poor vacuum in the film transport chamber which allows rapid pump down (<3 min) and fast access (<10 sec) for a 250 foot roll of film.

Since the EBR system operation is fully automatic and under computer control, no specialized training of the operator is required.

During the past 14 years, Image Graphics, Inc. (IGI) has made numerous advances in electron beam recorder (EBR) technology as a high performance, hard copy output device, capable of recording and plotting complex imagery, text and graphics on a variety of films and media such as: direct recording processless film; silver halide film - wet processed; dry silver halide film - heat processed; dielectric film - toner processed; and electron resists - chemically processed. The EBR has been used in the production of Aerial and Satellite Imagery; Geophysical Seismic Data; Engineering Documentation and Drawings; Computer Output Microfilm (COM); Cartographic and Graphic Arts products; Computer Generated Movies; and for Computer Assisted Data Storage and Retrieval.

The EBR is capable of recording both vector and text data from Computer Aided Design, Publishing and Line Art Systems and raster data from Image Scanners, Raster Image Processors (RIPS), Halftone/Screen Generators and Remote Image Sensors. A variety of image formats may be recorded on numerous film sizes (16mm, 35mm, 70mm, 105mm, 5", 5 1/2" and 9 1/2"). These recordings are used directly or optically enlarged depending upon the final product.

DATA COMPRESSION OF SPACEBORNE REMOTELY SENSED IMAGERY FROM COHERENT AND INCOHERENT SENSOR SYSTEMS

S. A. MANGO and K. W. HOPPEL

Radar Division
Naval Research Laboratory
Washington, D.C. 20375-5000
Tel: (202) 767-2003
FAX: (202) 767-5599
TELEX: 89 27 78

ABSTRACT

Some of the highest data rate, highest data volume systems of the past decade and projected for the next decade are spaceborne remote sensing imaging systems. These include the high-resolution, coherent synthetic aperture radar (SAR) systems as well as the high-resolution, incoherent visual and IR systems.

Considerations of partial or complete on-board processing, high speed data distribution to multiple users with diverse requirements and basic data archiving for these very demanding imaging systems have initiated or revived many data compression investigations.

This paper presents the results of:

1. a critical comparison of four effective data compression techniques applied to coherent SAR imagery and incoherent visual and IR imagery;
2. a study of the effectiveness of data compression applied to SAR data at various stages of processing: pre-, mid-, and post-processing and
3. a new variation of an adaptive discrete cosine transform technique to enhance low contrast and low activity areas in busy scenes.

1. The emphasis is on high-compression ratios in the range from 4-to-1 to 300-to-1 for the coherent SAR sensors, SEASAT and SIR-B and the incoherent visual/IR sensors LANDSAT TM and NOAA AVHRR.

2. The raw data from a coherent, SAR sensor system are more like a complex interferogram than an image. Each point source's energy in the scene domain is dispersed over a very large area in the raw data domain. In customary SAR processing a two-dimensional (range and azimuth), or more often, two sequential, one-dimensional match filtering process(es) collapse(s) the spread of energy from each point source into a two-dimensional radar image. The SAR processing steps provide opportunities for the application of data compression techniques at several stages whether the processing is to be performed on-board, on the ground or partially in both places. Data compression has been successfully applied to SAR raw data (pre-processing stage)[1], after range processing (mid-processing stage)[2] and several studies to image data (post-processing stage), e.g.[3]. However, conventional data compression techniques when applied to SAR raw data or range processed data have not produced comparable results to their application to SAR image data for the same compression ratio, apparently because the raw and range processed data are phase sensitive to changes in the reconstructed data.

This paper presents a critical comparison of near-optimum, lossy data compression techniques operating on pre-, mid- and post-processed SAR data over the large compression ratio range of 4-to-1 to 100-to-1. The techniques include

the spatial domain vector quantization (VQ) and block truncation coding (BTC) and the transform domain adaptive discrete cosine transform (ADCT) and adaptive Hadamard transform (AHT). The transform techniques, especially the ADCT, were found to be well-matched to the SAR specific speckle "noise" characteristics and the very large dynamic range at all stages of the SAR processing. Performance measures from both subjective polling of different viewers and quantitative measures such as the normalized mean-square-error (NMSE) and signal-to-noise ratio (SNR) were employed. However, the importance of applying application specific criteria as the "acid-test" for evaluating data compression results such as the ability to perform: 1.) feature classification, or 2.) change analysis instead of just 3.) simple detection of features is demonstrated with SAR imagery of sea ice.

3. In the Chen and Smith adaptive coding technique,[4] a DCT is performed for each subscene (block) of an image. The transformed blocks are sorted into classes by activity level which is measured by the block total ac energy. Within each class bits are allocated according to the variance matrix of the transformed data based upon an analytical optimization of the MSE. Bits are then distributed between higher activity levels (busy) and lower activity levels (quiet) to achieve the adaptivity. For the quiet classes this generally leads to fewer retained transform coefficients and a lower quantization (bits/transform coefficient) compared to the busy classes. Although the Chen and Smith method produces excellent results for several imagery types, it is naturally biased against low contrast areas embedded in busy or mixed activity scenes. This is the case in ocean and coastal scenes in which the body(ies) of water are often of low contrast and low relative activity but which may be the target areas of highest interest.

A new variation of an ADCT technique is described and illustrated using SEASAT, SIR-B, LANDSAT TM and NOAA AVHRR imagery. This technique forms the same activity classes as the Chen and Smith technique; however, it forces the number of retained transform coefficients to be uniform for certain or all the activity classes. The quantization levels of the bit allocation matrices absorb the major burden of the adaptive compression. The effect is to retain more of the high-frequency content in the low contrast, quiet areas so that subtle contrast changes in the original are usually better preserved. A small penalty, in some cases, of a slightly increased NMSE is shown to be offset by improved visibility of contrast differences in the quiet areas with little or no visible differences in the busy areas. The results of a parametric analysis and trade-off for the number of activity classes selected, the block size of the subscene and the compression ratio are presented.

REFERENCES

- [1] S. A. Mango, K. W. Hoppel, P. Bey, M. R. Grunes, "Case Study: Data Compression Techniques for SAR Data", Scientific Data Compression Workshop, Snowbird, Utah, May 1988.
- [2] P. LaFrance and J. Ousborne, "Effects of Reduced SAR Azimuth Processing," IEEE Trans. on Aerospace and Elec. Syst., AES-23, pp 152-157, March 1987.
- [3] C. Y. Chang, R. Kwok and J. C. Curlander, "Image Coding of SAR Imagery," Proc. IGARSS '87 Symposium, Ann Arbor, 18-21 May 1987.
- [4] W.-H. Chen and C. H. Smith, "Adaptive Coding of Monochrome and Color Images," IEEE Transactions on Communications, COM-25, pp 1285-1292, November 1977.

Data Compression Scheme for Archiving of ADEOS Data

Kohei Arai

Earth Observation Center,
National Space Development Agency of Japan

1401 Ohashi, Hatoyama, Saitama, 350-03 Japan

ABSTRACT

Some experiments have been conducted to investigate a potential ability on data compression of earth observation satellite data in particular, imaging sensor data. Two types of compression algorithms, information loss less and information lossy algorithms have been attempted. Compression ratios for both algorithms were 2.24 and 224, respectively, for SPOT/HRV data of Tokyo Japan.

1. INTRODUCTION

Advanced Earth Observing Satellite(ADEOS) will be launched with two core mission instruments and Announcement of Opportunity(AO) sensors in 1994 JFY. One of the core mission instruments, Advanced Visible and Near Infrared Radiometer(AVNIR) has capability to observe earth's surface with the spatial resolution of 8(m) for panchromatic band and 16(m) for multispectral bands consists of more than 5000 elements of CCD array results in increasing of the data rate of more than 120 Mbps. As for the ground station of ADEOS data acquisition, recording archiving, processing, etc, it is required to increase throughput, and archiving capacity.

Archiving of earth observation satellite data is getting more serious problem due to the fact that data rate of mission instruments onboard earth observation satellites is getting higher. Ground station operators have been investigating possibilities of new medias such as optical tape, optical disk, etc and of data compressions. One of active committee of data archiving, format and so on is Committee on Earth Observation Satellites(CEOS)/Working Group on Data(WGD). In the committee although new medias have been investigating, not so many data compression algorithms have been studied

so far that some experiments have been conducted.

Data compression algorithms can be divided into two categories, Information Loss Less and Information Lossy, in other words, Reversible and Irreversible algorithms. As for raw imagery data, original information should be preserved so that information loss less algorithm is one and only one candidate while for quick look data, not only information loss less, but also information lossy algorithms are available because image quality degradation is permitted for utilization of quick look imagery data.

Firstly data compression algorithms will be overviewed followed by some descriptions for the reversible predictive coding(information loss less) and the irreversible predictive coding(information lossy). Second results from the experiments with SPOT/HRV data of Tokyo Japan will be described with the detailed information on parameters of the experiments such as statistical properties of the original SPOT/HRV data used, decimation factor, and so on together with data compression ratios for a variety of parameters.

2. Data compression algorithms

2.1 Information loss less

Fundamental compression techniques on the information loss less coding is Entropy(Huffman) coding(Ref.1). The Huffman coding, however, is not so efficient in terms of compression ratio that the following algorithms have been proposed. For multilevel imagery data, Rice coding(Ref.2), Bit plane coding(Ref.3), etc have been proposed while for natural images with many quantization levels, Differential entropy coding, Block adaptive reversible predictive coding(Ref.4), etc have also been proposed. As for data compression algorithms for earth observation satellites data, later is

sueteable. In terms of data compression ratio, in general, block adaptive reversible predictive coding is superior to differential entropy coding. Thus the block adaptive reversible predictive coding is attempted.

2.2 Information lossy

Information lossy algorithms can be divided into three major categories, Predictive coding, Transformation coding and Quantization (in particular, Vector quantization). The predictive coding contains Differential pulse code modulation (DPCM) (Ref.5), etc while the transformation coding includes Karhunen-Loeve transformation (Ref.6), Fourier transformation (Ref.7), Walsh-Hadamard transformation (Ref.8), Slant transformation (Ref.9), Cosine transformation (Ref.10), Pyramid structural Hadamard transformation (Ref.11), etc. On the other hand, Vector quantization coding has a variety of combination with cosine transformation (Ref.12), quantization speed accelerated coding (Ref.13), and so on.

Data compression ratio, in general, depends upon acceptable image quality gradation, statistical property of original imagery data, etc. In this study commonality with information loss less algorithm, reversible predictive coding, irreversible predictive coding has been attempted.

2.3 The reversible predictive coding

Algorithm of the reversible predictive coding is as follows,

- (1) Partitioning an original image with 8x8 pixels blockes.
- (2) Calculation of prediction error: e of each pixel in the block of interest by means of the following equation.

$$e = 2X - (X1 + X2) \quad (1)$$

where X denotes pixel value of interest, $X1$ is the pixel value of just one pixel ahead of the pixel of interest while $X2$ is the pixel value of just one pixel above the pixel of interest.

- (3) Shrinking of the prediction errors by means of the following equation.

$$f = \begin{cases} 1 & \text{if } e = 0 \\ g + 1 & \text{if } 0 < g < 255 \\ 512 - g & \text{if } 255 < g < 511 \end{cases} \quad (2)$$

$$g = \begin{cases} e - 1 & \text{if } e > 0 \\ 511 + e & \text{if } e < 0 \end{cases} \quad (3)$$

- (4) Derive an encoding parameter i

$$i = \begin{cases} 0 & \text{if } S(f) < [2.62 \times 64] \\ 1 & \text{if } [2.62 \times 64] < S(f) < [4.68 \times 64] \\ 2 & \text{if } [4.68 \times 64] < S(f) < [8.82 \times 64] \\ 3 & \text{if } [8.82 \times 64] < S(f) < [17.13 \times 64] \\ 4 & \text{if } [17.13 \times 64] < S(f) < [33.73 \times 64] \\ 5 & \text{if } [33.73 \times 64] < S(f) < [65.15 \times 64] \\ 6 & \text{if } [65.15 \times 64] < S(f) < [105.00 \times 64] \\ 7 & \text{if } [105.00 \times 64] < S(f) \end{cases} \quad (4)$$

where $S(f)$ is summation of g within a block while $[y]$ denotes Gauss symbol, an integer part of y .

- (5) Encoding f with the following equation.

$$\begin{array}{c} \begin{array}{c} i \\ \text{[}(f-1)/2 \text{] bit} \\ \text{-----} \\ f \rightarrow 1 \ 1 \dots\dots\dots 1 \end{array} \\ \begin{array}{c} j \text{ bit (binary expression of } h) \\ \text{-----} \\ x \ x \dots\dots\dots x \end{array} \end{array} \quad (5)$$

$$\text{where } h = (f-1) - \left[\frac{(f-1)}{2} \right] 2^i \quad (6)$$

$$j = \begin{cases} i & \text{if } \left[\frac{(f-1)}{2} \right] + 1 \neq 256 \\ i + 1 & \text{else} \end{cases} \quad (7)$$

2.4 The irreversible predictive coding

The algorithm of the irreversible predictive coding is as follows,

- (1) Partitioning, calculation of prediction error and shrinking of the prediction error are exactly same as the aforementioned algorithm.
- (2) Quantization of the shrunk prediction error into 16 levels by means of the following equation.

$$q = \left[\frac{f}{2} + 0.5 \right] \quad \text{if } q > 16, \text{ then } q = 16. \quad (8)$$

- (3) Huffman coding is applied to the quantized data q .

3. Results from experiments

3.1 SPOT/HRV data used

SPOT/HRV panchromatic band data of Tokyo Japan is used. Small portion of the image used is indicated in Fig.1 while description of the image used is shown in Table 1.

3.2 Statistics of prediction error

As for data compression of quick look images, it is still useful if decimations is applied to the original imagery data so that statistics for prediction errors, firstly, is studied.

The statistics with the range of decimation factor from 1 to 10, are shown in Table 2 and are indicated in Fig.2.

Table 1 Description of SPOT/HRV panchromatic band data used

| Item | Description |
|---------------------------------|---|
| Acquisition date | Oct.17 '86 |
| Scene center time | 01:35:11(UT) |
| Scene center location | 35:23:07(N) - 139:37:37(E) |
| Grid reference | 331 - 279 |
| Image size | 6316 pixels by 6004 lines |
| Image size used for experiments | 1200 pixels by 1200 lines (Top left corner of the image) |



Fig.1 A portion of the SPOT/HRV panchromatic band data used for the experiments on data compression.

Table 2 Statistics of prediction error

| prediction error | decimation factor | 1 | 2 | 3 | 5 | 10 |
|------------------|-------------------|------|------|------|------|------|
| 0 | | .177 | .114 | .092 | .071 | .061 |
| 1 | | .274 | .194 | .164 | .138 | .115 |
| 2 | | .172 | .144 | .130 | .115 | .102 |
| 3 | | .115 | .110 | .103 | .099 | .090 |
| 4 | | .080 | .086 | .086 | .084 | .081 |
| 5 | | .056 | .070 | .072 | .072 | .071 |
| 6 | | .038 | .057 | .061 | .064 | .064 |
| 7 | | .026 | .046 | .051 | .054 | .058 |
| 8 | | .018 | .036 | .043 | .047 | .049 |
| 9 | | .012 | .029 | .036 | .041 | .045 |
| 10 | | .008 | .023 | .029 | .035 | .040 |
| 11 | | .006 | .018 | .025 | .030 | .033 |
| 12 | | .004 | .014 | .020 | .026 | .027 |
| 13 | | .003 | .011 | .016 | .022 | .026 |
| 14 | | .002 | .009 | .013 | .016 | .024 |
| 15 | | .002 | .007 | .010 | .015 | .018 |
| 16 | | .001 | .006 | .009 | .012 | .015 |
| . | | | | | | |
| average | | 35.0 | 34.8 | 35.0 | 35.0 | 34.8 |
| variance | | 75.0 | 76.2 | 79.1 | 79.6 | 79.9 |

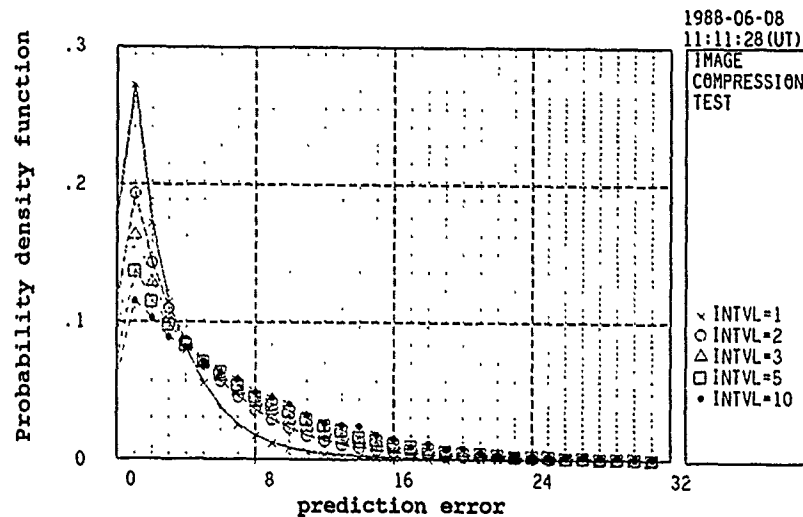


Fig.2 Probability density function of prediction error

3.3 Data compression ratio

3.3.1 The reversible predictive coding

In terms of data compression ratio Cr defined as the following equation, 2.24 can be achieved for the imagery data used.

$$Cr = n \times m \times b / c \quad (9)$$

where n, m are image size(n pixel by m line), b is quantization bit, and c is the number of coding bit.

3.3.2 The irreversible predictive coding

Data compression ratios of the coding algorithm with the parameter of the decimation factors from 1 to 10 are shown in Fig.3 and Table 3.

4. Concluding remarks

As for quick look imagery data catalogue, image database is required. In order to save the space for the database, not only new mass medias with high access speed, but also data compression is taken into account. For this purpose, information lossy algorithms are applicable. The experimental results show that 224 of data compression ratio can be achieved with the decimation factor of 10, for SPOT/HRV panchromatic data by means of

the irreversible predictive coding.

Meanwhile information loss less coding algorithms are applicable for space saving of archiving of raw and processed data. The experimental results also show that 2.24 of data compression ratio can be achieved for the same data by mean of the reversible predictive coding.

Further investigation is required for a variety of algorithms and imagery data.

References

- (1) P.A.Wintz : Transform picture coding, IEEE Proceedings, vol.60, no.7 pp.809 - 820, 1972.
- (2) R.F.Rice and J.R.Plaunt : Adaptive variable-length coding for efficient compression of spacecraft television data, IEEE Trans. COM - 19, 6, pp.889 - 897, 1971.
- (3) J.W.Schwartz and B.C.Barker : Bit-plane encoding, A technique for source encoding, IEEE Trans., AES - 2, 4, pp.385 - 392, 1966.
- (4) W.K.Pratt : Digital Image Processing, Wiley-Interscience, 1978.
- (5) T.S.Huang(ed.) : Digital Picture Processing, Academic Press, 1982.
- (6) G.S.Robinson : Orthogonal Transform Feasibility Study, NASA Final Report, NASA-CR-115314, N72-13143, pp.176, 1971.

Table 3 Data compression ratio of the irreversible predictive coding

| Decimation factor | 1 | 2 | 3 | 5 | 10 |
|-------------------|------|-------|-------|-------|--------|
| Compression ratio | 3.70 | 10.91 | 22.48 | 58.62 | 223.97 |

(7) P.A.Wintz : Transform Picture Coding, IEEE Proceedings, vol.60, no.7, pp. 809 - 820, 1972.

(8) G.S.Robinson : Quantization Noise Considerations in Walsh Transform Image Processing, Proceedings of the 1972 Symposium on Applications of Walsh Functions, pp.240 - 247, 1972.

(9) W.H.Chen and W.K.Pratt : Color Image Coding with the Slant Transform, Proceedings of the 1973 Symposium on Applications of Walsh Functions, 1973.

(10) N.Ahmed, T.Natarajan and K.R.Rao : Discrete Cosine Transform, IEEE Trans. on COM, 1977.

(11) Y.Tasuda, M.Takagi, S.Katou and T.Awano : Transmission and display of still picture data using pyramid structured Hadamard transform, Elec. Comm. Journal, Trans. vol.63 - B, no.4, 1980.

(12) Saitou, Takeo, Harashima and Miyakawa : Gain/Shape Vector Quantization for multidimensional signal with spherical and symmetric distribution, ibid, J68 - B, pp.904 - 911, 1985.

(13) R.M.Gray : Vector Quantization, IEEE ASSP Magazine, 4 - 29, 1984.

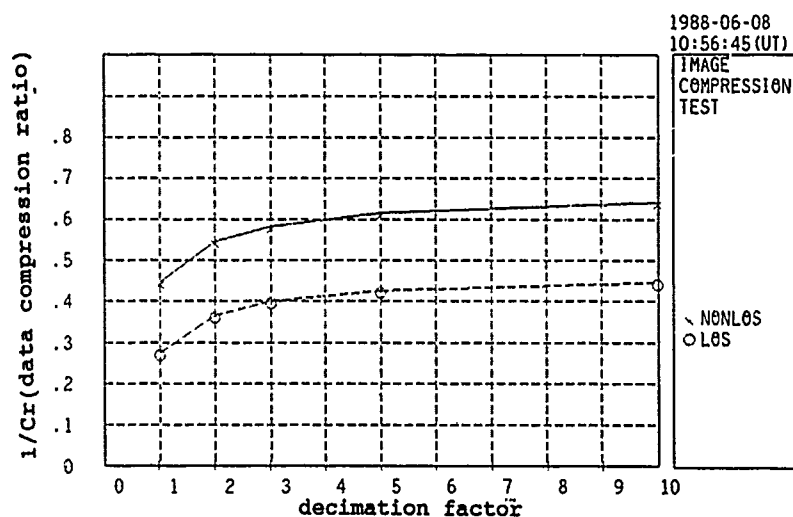


Fig.3 $1/Cr$ (data compression ratio) as the function of decimation factor

- ABSTRACT -

NOISELESS IMAGE COMPRESSION

Kristo Miettinen
and
Dr. Nicholas Baser

General Electric
M2412
230 Goddard Blvd.
King of Prussia, PA
U.S.A. 19406

Phone: (215) 354-5713
(215) 354-2814

FAX: (215) 354-2226

Noiseless image compression is an essential technique for reducing transmission bandwidth and data storage requirements for sensor systems requiring total fidelity to the original sensor data. An effective technique for noiseless compression of all image data has been developed from the techniques applied to the Voyager II Uranus encounter.

The algorithms developed and published by Robert F. Rice of Jet Propulsion Laboratories (hereafter the JPL algorithms) for Voyager II image compression rely on encoding small image blocks with members of a family of simple predefined codes. Because the codes are not developed ad hoc at the encoder (as are huffman codes) they are necessarily not optimum for the individual blocks being encoded, but because the most applicable coder can be selected for each block from the family available, each block will be reasonably efficiently encoded. Furthermore, since there is little overhead involved in changing codes, the image blocks processed with a coder can be very small, allowing rapid changes in coder choice to match varying image characteristics. This ability to change coders frequently (as often as every eight or sixteen pixels) enhances performance so much that the overall compression result is usually a data set smaller than the inherent entropy of the image taken as one block. For example, an image of a simulated mars landing site with an image entropy of 6.98 bits per pixel and a differential image entropy of 4.00 bits per pixel was compressed to 3.72 bits per pixel with a software implementation of the current algorithm.

The current algorithm is a refinement of the JPL algorithm set based on research for a proposed mars sampling rover mission. It involves a family of encoders with effective operating points at integer entropies: one coder operates well on data sets with no entropy (each difference image pixel must be zero), one operates well on blocks with entropies of one bit per pixel, one on blocks with entropies of two bits per pixel, and so on up to the dynamic range of the sensor (a sensor producing a stream of eight-bit numbers cannot produce a data set with entropy exceeding eight bits per pixel). The encoders are applied to four pixel by four pixel blocks of a difference image, and the map of encoders used is further compressed and included as overhead.

Effective, efficient and simple noiseless image compression has implications beyond noiseless compression, since even transform-based compression schemes leading to inexact image reconstruction involve exact (noiseless) encoding of quantized transform coefficients. Effective inexact-reconstruction compression schemes without orthonormal basis transformations have also been developed from the noiseless encoding algorithm.

The algorithm developed for the proposed Mars Rover Sample Return mission has been designed specifically with simple, low-cost, low-power hardware implementation in mind. A proposed AVLSI implementation would require 40,000 transistors for logic functions and eleven seperate FIFO memories for pixel line buffers.

ABSTRACT
DATA COMPRESSION ON THE MARS ROVER
SAMPLE RETURN MISSION

Dr. Nicholas Beser
and
Kristo Miettinen

General Electric
M2412
230 Goddard Blvd.
King of Prussia, PA
U.S.A. 19406

Phone: (215) 354-2814 or (215) 354-5713

FAX: (215) 354-2226

The Mars Rover Sample Return mission will orbit an observation platform around Mars, survey potential landing sites, and determine the expected terrain for both landings and robotic exploration of the planet. A cornerstone of the mission is the ability to predict the nature of the target area by remote sensing means. The resolution of the sensor data will surpass any previous interplanetary probe. Projected resolution could achieve 0.25 meters per pixel. The high resolution coupled with potential orbital configurations indicate a very high data rate from the sensor. The data rates, and potential data volume of the target scene suggest that either an extremely large on board memory will be required, or a data compression system will be used to reduce the storage requirements. In addition, the data communications channel between Mars and Earth will be an effective 100KBits per second when taking into account error correcting codes. This indicates that the time period for transmission is too short to send the uncompressed data.

Data compression has been used in this study as a focal point for investigating many of the system issues involved with trades between the sensor platform, communications, data handling subsystem, and user applications. The impact of sensor configurations, data collection scenarios,

and orbital configurations have been traded against compression hardware designs. The effect of communication errors on image reconstruction as well as protocol control of the data handling subsystem has been investigated. Finally, different compression algorithms have been studied for potential flight implementation. The study included simulated Mars landing scene compression and reconstruction to determine the effect on image quality. Data compression algorithms included Variable Bit Rate Zonal Compression, Scene Adaptive Compression, and Noiseless Encoding Methods.

IHS TRANSFORM FOR THE INTEGRATION OF RADAR IMAGERY WITH GEOPHYSICAL DATA

INTERA Technologies Ltd./CCRS

J. Harris and R. Murray

ABSTRACT

The IHS (intensity-hue-saturation) transform is a powerful technique for the integration of digital data. This paper briefly reviews the theory behind the IHS display technique and illustrates how it can be applied for integrating radar and geophysical data. The resulting IHS transformed colour images show textural/topographic information provided by the radar as a function of image intensity, combined with hue (colour) information provided by the geophysical data. The transformed images provide a wealth of geological information and are suitable for both qualitative and quantitative analysis.

KEYWORDS: IHS, data integration, radar, geophysics

1.0 INTRODUCTION

Concomitant with the increase in computer image analysis and Geographic Information System (GIS) technology, more emphasis today is being placed on the digital integration of diverse data types (Aarnisalo et al., 1982; Haydn et al., 1982). Data integration is obviously not a new concept and has been pursued for many years on an analog basis in many earth science disciplines. However, rapid advances in image analysis hardware and software have allowed for greater flexibility and new techniques for combining and integrating digital data.

Commonly remotely sensed data such as Landsat MSS is displayed on a video monitor using the additive primaries red, green and blue (RGB) or recorded on a plotting device through the mixture of the subtractive primaries yellow, cyan and magenta. The RGB system is intuitively simple and, in many cases, effective for displaying a maximum of three image channels. However, it is often difficult to perceive subtle differences in colour using this system and a quantitative description of colour is difficult to achieve.

A more appropriate colour display system would allow for the linear mapping of the data to orthogonal variables. A colour system that allows for this type of mapping is the IHS colour transform. The IHS system has been much discussed (Pratt, 1978; King et al., 1984; Gillespie, 1980; Buchanan, 1979) and various applications for the display of remotely sensed data using this colour display transform have been demonstrated (Haydn et al., 1982; Daily, 1983; Raines, 1977; Kruse and Raines, 1984; Gillespie et al., 1986; Robertson and O'Callaghan, 1988). This paper describes how the IHS transform can be used as a technique for integrating radar with airborne geophysical data (magnetics and gamma ray spectrometer) providing

imagery in which image colour (hue) can be interpreted in both a relative and absolute sense. Radar imagery is used as a base product for integration for a number of reasons. Much emphasis is being placed on radar as an effective tool for earth sensing and observation as many countries, including Canada (RADARSAT), US (SIR-C), Europe (ERS-1) and Japan (JERS-1), are now actively involved in the development of space borne radar systems. Radar, because of its side viewing geometry and longer wavelengths which results in an all weather sensing capability, has established itself as an extremely effective sensor for earth observation, as a unique view of the terrain can be achieved, making it useful for a variety of geoscience studies in which information regarding terrain geometry (topography), surface roughness and moisture content are important variables. The enhancement of terrain morphology is critical for geological applications as topographic patterns often reflect underlying geologic structure and lithology.

2.0 METHODOLOGY

A plethora of colour co-ordinate systems have been developed over the past 40 years, most of which have been developed to quantify colour photographs and predict human perception (Gillespie, 1980). Although the RGB colour system, commonly used to display three channel remotely sensed imagery, is simple and often effective, a number of shortcomings exist as summarized by Robertson and O'Callaghan (1988). The RGB system is not based on readily definable colour attributes and therefore, colour variations as defined by the mix of red, green and blue primaries are not always easy to perceive and/or to describe numerically. This results in displays in which the numerical characteristics of the data are not represented by uniform colour gradations.

An effective display co-ordinate system which can overcome many of these shortcomings is the IHS transform, which is defined by three separate, orthogonal and easily perceived colour attributes, that of intensity, hue and saturation. Geometrically, the RGB system can be represented as a cube with the red, green and blue axes defining the x, y and z vectors respectively. The IHS co-ordinate system can be represented as a cylinder or a sphere. Intensity, which represents the total energy or brightness of the image, defines the vertical axis of the cylinder, or the radius of the sphere, while hue, which represents the average wavelength of colour, defines the circumferential angle of the cylinder or sphere, and ranges from blue (0 degrees) through green, yellow, red and purple (360 degrees). Saturation can be thought of as the purity of the colour (i.e. percentage of white light in the image) and defines

the colatitude of the sphere, or the radius of the cylinder. The mathematics involved in the transform from cartesian (RGB) to spherical or cylindrical (IHS) co-ordinates are reviewed by Gilliespie (1980) and King et al. (1984), while Haydn et al. (1982) provide a general descriptive review of the IHS system.

The advantages of the IHS co-ordinate system over the RGB system are that the informative aspects of an image are presented using readily identifiable and quantifiable colour attributes that can be distinctly perceived. Secondly, numerical variations in the image data can be uniformly represented in an easily perceived range of colours and, thirdly, individual control over the chromatic (hue) and achromatic (saturation) components of the image is possible. Furthermore, mapping different data types into the IHS colour space can produce more complex images in which variables with diverse information content can be represented by different colour attributes. It is also possible to produce images which are a combination of more than three channels, thus providing more information in the resultant colour composite image after transformation back to RGB space for display on a video monitor.

Several hardware and software components were employed to create the images and perform the analysis described below. A computer image analysis system and associated software, available from Dipix Technologies Ltd. (ARIES-III), the Film Image Recorder (FIRE) from MacDonald Dettwiler and Associates, and software written to perform colour space conversions written by Intera Technologies Ltd. were all used to generate the examples of colour composite imagery shown in this paper.

The software used three related IHS type transformations, one based on a spherical mathematical model and two based on cylindrical transformations. The transformations are discussed in the full version of this paper which is presently in press (Harris et al., 1989).

Two examples of integrated radar and airborne geophysical data of eastern Nova Scotia are shown and their value for geological mapping activities are summarized. Figure 1 is a flow diagram summarizing the steps required to produce the IHS transformed imagery discussed in this paper.

3.0 INTEGRATION OF RADAR AND GEOPHYSICAL DATA

The IHS transform is an effective technique for integrating radar data with diverse data types such as airborne geophysical data, providing imagery which displays a unique and often very informative "picture" of the earth's surface and, in the case of airborne magnetics, subsurface characteristics. Particularly important when producing these IHS combined images is the assignment of hue, as this will provide image colour which can be directly related to an absolute measurement scale.

Figure 2 is an IHS transformed colour composite image (shown in black and white - see Harris, 1989 for colour version) which is a combination of an airborne radar image and a single channel digital vertical gradient magnetics image over eastern Nova Scotia. The C-band radar data was acquired by the Canada Centre for Remote Sensing (CCRS) in wide swath mode spanning a range of incidence angles from 45 to 80 degrees. Further details on the CCRS C-band radar system are given in Livingstone et al. (1987). The airborne magnetic data was acquired digitally, compiled and gridded by the Geological Survey of Canada (Hood, 1979). The radar image was registered to a Universal Transverse Mercator (UTM) geometrically correct topographic map base and reformatted from 12.5 m pixels to 50 m pixels. The magnetic data was co-

registered to the radar data using UTM co-ordinates as control points.

The eight bit vertical gradient magnetic digital image with values ranging from 0 to 255 DN, was density sliced into 12 discrete levels representing absolute measurements of the magnetic vertical gradient in units of gamma (γ) and this density sliced image was used to modulate image hue. These 12 levels were mapped into the hue spectrum so that low levels of gamma (negative values) are represented in blue and green while higher levels (positive values) range from orange through to red and purple (see Figure 2). Since the minimum and maximum gamma values were mapped to 0 and 255 DN respectively, the density slices and, subsequently, image hues could be absolutely calibrated to units of gamma. The radar data was used to modulate image intensity while a saturation file was synthetically generated and assigned a DN level of 150 to provide hues that were less vibrant, thus ensuring a proportionate mix of the radar and magnetic data. A final step involved creating a saturation mask for the gamma values ranging from -1 to +1, as the data in this range provided little structural information. This saturation mask was set to zero to ensure that these areas would be black and white on the final image. Thus, only the linear structural anomalies present on the vertical gradient data are highlighted in colour. These three IHS channels were then reverse transformed to RGB space to produce the viewable image.

This IHS transformed image provides a useful product for geologic exploration purposes on a number of levels. Firstly, the cartographic information, such as lakes, roads and urban areas, provided by the radar helps to better locate and evaluate the patterns present on the magnetic data. This can be especially important when undertaking field programmes. Secondly, this image provides a unique product for both qualitative and quantitative photogeologic analysis as the colour gradations represent calibrated gradient levels. The combination of surface topographic patterns supplied by the radar and magnetic anomalies reflecting near surface structural and lithologic features results in an image which effectively portrays regional geologic structure. Furthermore, verification and correlation of interpreted geologic features is facilitated as the image contains information from more than one geologic data source. Mapping of regional structure is particularly important in this area as an aid to on-going gold exploration activities.

Figure 2 represents a radar/magnetic IHS transformed image in which a single channel (magnetics) has been used to provide colour information. However, multiple channels may be used in the IHS transform to provide colour information as suggested by Buchanan (1979). Figure 3 is an example of a radar/gamma ray spectrometer IHS colour image (shown in black and white - see Harris et al., 1989 for colour version) covering an area in eastern Nova Scotia, Canada in which the hue information has been supplied by three gamma ray spectrometer channels, equivalent uranium (eU), equivalent thorium (eTh) and percent potassium (%K). A C-band wide swath radar image is used to modulate image intensity. The airborne gamma ray spectrometer data was acquired digitally and compiled and gridded to 200 metre pixels by the Geological Survey of Canada (Grasty, 1972). The data was then resampled to 50 metre pixels and registered to a UTM topographic base. The radar image was acquired by CCRS at a pixel size of 12.5 metres. The image was subsequently resampled to 50 metre pixels and co-registered with the UTM topographic map base.

The image production process outlined in Figure 1 consisted of equalizing the means and standard deviations of each of the three spectrometer channels and stretching the minimum and maximum values to cover the full range of the

eight bit data (i.e. 0 to 255 DN). The three spectrometer channels were then input into the IHS transform and the radar data were used to replace the intensity channel before converting back to RGB space. The colour triangle associated with Figure 3 provides a colour guide with which to interpret the relative mix of the eU, eTh and %K channels. The particular channel, as well as the maximum values for each spectrometer channel (ie ppm for eU and eTh and % for K), are shown at the corners of the triangle. Areas high in eU are red, high in eTh are green and high in %K are blue. Proportionate mixes of the primary colours result in magenta, cyan and yellow colours that can be interpreted on a relative basis as mixtures of the three spectrometer channels. Thus, yellow areas have roughly equal proportions of eU (red) and eTh (green), while cyan areas have comparable proportions of eTh (green) and %K (blue). Saturation in this image was derived from the original RGB to IHS transformation (see Harris et al., 1989 for mathematical formulation of the three spectrometer channels).

This IHS transformed image represents a multichannel colour composite image as it is a combination of four data channels (radar + eU, eTh, %K) as opposed to a standard three colour composite. This experimental IHS image has been used to aid in the mapping of lithology, particularly granites, and regional structural patterns in eastern Nova Scotia (Harris, 1989). The data types comprising the imagery are complementary, with the radar providing a map of the terrain surface in which topographic patterns are enhanced due to the side viewing geometry and the spectrometer data providing a picture of the radiometric "landscape". The two different and unique views of the terrain contained in one image facilitate photogeologic interpretations, as interpreted features can be compared and more easily verified from a geological perspective.

Further details on the use of these IHS transformed images (Figs. 2 & 3) for geological mapping can be found in Harris, 1989 and Harris et al., 1989.

4.0 SUMMARY AND CONCLUSIONS

A methodology for creating experimental colour image products combining airborne radar with airborne geophysical data using the IHS colour display transform has been demonstrated. Although this methodology is applicable for the integration of virtually any digital data set, radar has been used as the base product for integration, as it provides a good cartographic base in which topographic, morphologic and surface textural patterns are enhanced. This is particularly important when integrating with geophysical data, as the radar provides a recognizable map of the terrain surface, facilitating a comparison between terrain and geophysical patterns.

The IHS colour display transform is a useful method for the integration and unambiguous and controlled portrayal of diverse digital data sets. Greater control over the image construction process is possible as individual data channels can be assigned to the quantifiable and easily perceived colour parameters of intensity, hue and saturation. By controlling image hue, the association of a meaningful colour scheme with well defined characteristics of the input data can be achieved. The image hues can be interpreted on a relative or absolute basis, depending on what and how the data was mapped to the hue parameter. Furthermore, single and multiple channels can be used to supply information on image hues as demonstrated by the image examples discussed in this paper.

ACKNOWLEDGEMENTS

The author would like to thank the internal CCRS reviewers of this paper, particularly V.R. Slaney, for constructive comments. Blair Moxon's help with the design and computer drafting of the Figures was invaluable. Airborne radar data was provided by Intera Technologies Ltd. and the Canada Centre for Remote Sensing (CCRS), while the airborne geophysical data was supplied by the Geological Survey of Canada (GSC). This work was carried out under CCRS contract OSIN:23413-7-9001, "Scientific and Technical Support for Radarsat".

REFERENCES

- AARNISALO, J., E. Franssila, J. Eeronheimo, E. Lakanen and E. Pehkonen, (1982), On the Integrated Use of Landsat, Geophysical and Other Data in Exploration in the Baltic, Sheild, Finland, Photogrammetric Journal of Finland, Vol. 9, no.1, p 48-64
- BUCHANAN, M.D., (1979), Effective Utilization of Color in Multidimensional Data Presentations, SPIE Vol. 199, Advances in Display Technology, p 9-19
- DAILY, M., (1983), Hue-Saturation-Intensity Split-Spectrum Processing of Seasat Radar Imagery, Photogrammetric Engineering and Remote Sensing, Vol 49, No. 3, March, p. 349-355.
- GILLESPIE, A.B., A.B. Kahle, R.E. Walker, (1986), Color Enhancement of Highly Correlated Images, I. Decorrelation and HSI Contrast Stretches, Remote Sensing of Environment Vol. 20, p.209-235.
- GILLESPIE, A.R., (1980), Digital Techniques of Image Enhancement in Remote Sensing in Geology, B.S. Siegal & A.R. Gillespie, Eds, New York, Wiley, 1980, p. 139-226.
- GRASTY, R.L., (1972), Airborne Gamma Ray Spectrometry Data Processing Manual, Geological Survey of Canada, G.S.C. Open File Report 109.
- HARRIS, J.R (1989), Analysis of Airborne Radar of Eastern Nova Scotia for the Mapping of Regional Geologic Structure: Implications for Gold Exploration (in press)
- HARRIS, J.R., R. Murray and T. Hirose (1989), IHS Transform for the Integration of Radar and other Remotely Sensed Imagery (in press)
- HAYDN, R., G.W. Dalke, J. Henkel, (1982), Application of the IHS Color Transform to the Processing of Multisensor Data and Image Enhancement, Presented at the International Symposium on Remote Sensing of Arid and Semi-Arid Lands, Cairo, Egypt, January, p 599-616
- HOOD, P.J., (1979), Magnetic Methods Applied to Base Metal Exploration, Geophysics and Geochemistry in the Search for Metallic Ores, P.J. Hood ed, Geological Survey of Canada, Economic Geology Report 31, p 77-104
- KING, R.W., V.H. Kaupp & W.P. Waite, (1984), Digital Color Space Transformations, Proceedings of IGARSS'84 Symposium, Strasbourg 27-30 Aug., p 649-654
- KRUSE, F.A., and G.L. Raines, (1984), A Technique for Enhancing Digital Color Images by Contrast Stretching in

Munsell Color Space, Proceedings of the International Symposium on Remote Sensing of Environment, 3rd Thematic Conference, Environmental Research Institute of Michigan, Colorado Springs, CO, p. 755-760

LIVINGSTONE, C.E., A.L. Gray, R.K. Hawkins, R.B. Olsen, J.G. Helbertsma, R.A. Dean (1987), CCRS C-Band Airborne Radar -System Description and Test Results, Proceedings of the 11th Canadian Symposium on Remote Sensing, Waterloo, June, p 503-518

PRATT, W.K., Digital Image Processing, A Wiley-Interscience Publication, John Wiley & Sons.

RAINES, G.L., (1977), Digital Color Analysis of Color-Ratio Composite Landsat Scenes, Proceedings of the International Symposium on Remote Sensing Environment, ERIM, Vol 11, p 1463-1477

ROBERTSON, P.K. and J.F. O'Callaghan, (1988), The Application of Perceptual Color Spaces to the Display of Remotely Sensed Imagery, IEEE Transactions on Geoscience and Remote Sensing, Vol. 26, No. 1, Jan. p 49-59

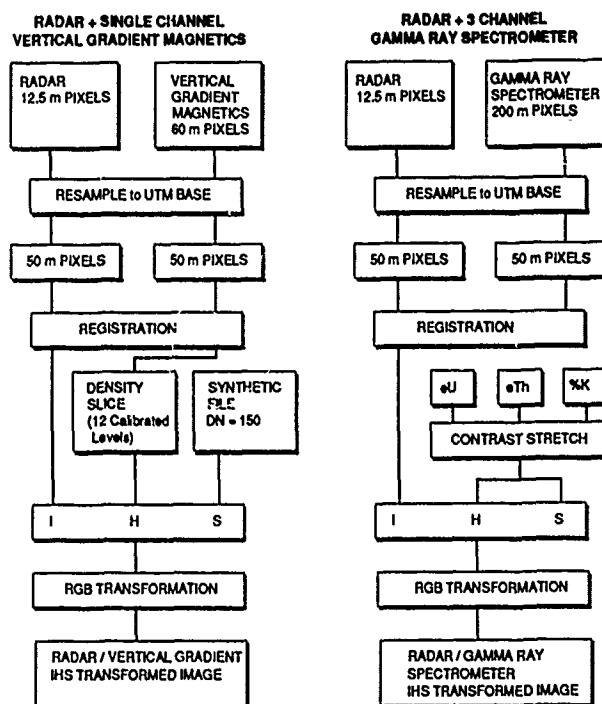


Figure 1 - Image Processing Methodology

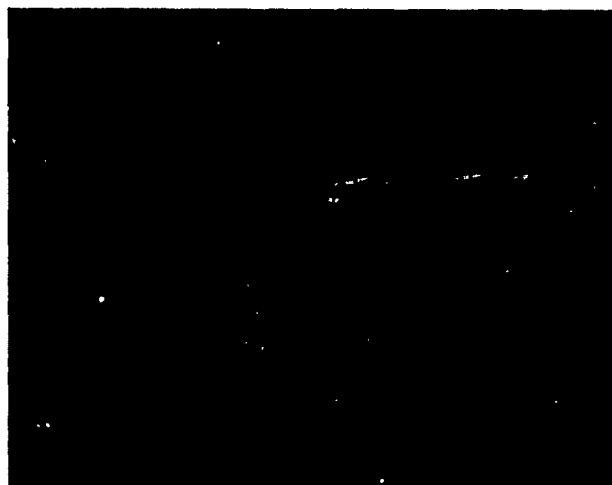
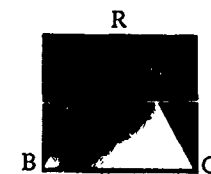
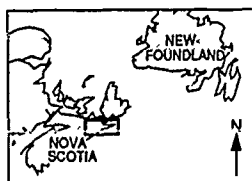


Figure 2 - Radar / Vertical Gradient Magnetism IHS Transformed Image



Figure 3 - Radar / Gamma Ray Spectrometer IHS Transformed Image



R = eU (max = 7 ppm)
G = eTh (max = 11 ppm)
B = %K (max = 2.82 %)

Figs. 2 & 3 are half-tone black and white reproductions of colour images, therefore, many of the image details have been lost. In Fig.3 red areas (high eU) are displayed in dark tones while green areas (high eTh) are displayed in light tones. The original colour images are to be published in upcoming papers by Harris et al. and are also available for viewing through CCRS.

POLARIZATION DIVERSITY RADARS IN METEOROLOGY

James I. Metcalf
Air Force Geophysics Laboratory
Hanscom Air Force Base, Mass., U. S. A.

ABSTRACT

Polarization diversity radar techniques can contribute to a variety of meteorological research topics and applications: the evolution of hydrometeors, the development of precipitation, cloud electrical phenomena, quantitative measurement of rainfall, discrimination of ice and liquid water in clouds, the detection of hail and of aircraft icing conditions, and the description of microwave and millimeter-wave propagation in clouds and precipitation. This broad spectrum of potential utility results from the capability of such radars to measure quantities related to the shapes and orientations of the hydrometeors that constitute the propagation and backscatter media. Through analytical or empirical models, estimates of particle sizes, shapes, orientations, and thermodynamic phases can be derived from the measured quantities. There has been significant recent progress both in the development of polarization diversity meteorological radar systems and in the interpretation of the resulting measurements.

Key issues that must be faced by radar meteorologists in the near future include (1) the more effective use of existing radars in research, both in multi-agency field programs and in locally managed operations; (2) improved verification of radar measurements, based on comparative measurements in the laboratory and in the field; and (3) technological improvement of meteorological radars by the incorporation of improved antennas, polarization control, receivers, and data processors. There is also a need to identify and pursue lines of research that could lead to the development of operational techniques. Expanded interactions between the polarization diversity radar specialty and the disciplines of cloud and precipitation physics, mesoscale meteorology, and hydrology are fundamental to progress in these areas.

THE DIFFERENTIAL REFLECTIVITY DUAL-POLARIZATION RADAR
TECHNIQUE: ACCOMPLISHMENTS AND FUTURE PROSPECTS IN
METEOROLOGY

Thomas A. Seliga
Communications and Space Sciences Laboratory
The Pennsylvania State University
University Park, PA 16802
TEL: 814-865-4542 FAX: 814-863-4749

This paper presents a review of progress on the use of the differential reflectivity radar technique in meteorology since its inception in 1976 by Seliga and Bringi (J Appl. Meteor., 15, 69-76, 1976). Following a brief two to three year period of experimental testing and evaluation, the concept has now become generally accepted as an excellent tool for rainfall measurement and the detection of ice phase hydrometeors. As suggested by theory, excellent comparisons between rainfall rates inferred from dual linear polarization radar reflectivity measurements and ground-based observations have been demonstrated due to the additional information available on raindrop size which is inferred from the differential reflectivity radar parameter. Early observations clearly indicated that the technique was also valuable for discriminating between rain and ice phase, and this led to its being acknowledged as an excellent means for detecting hail. The advances achieved to date, the relative simplicity of the measurements and their interpretation and the importance of the inferences derived from the measurements have combined to encourage many research groups to either add this polarization capability to existing research radars or to construct new radars based largely on the promises of the technique for cloud physics and related studies. Perhaps more importantly, the successes achieved to date have not only opened new fields for investigation but have also led the way for scientists to re-examine more closely other radar parameters in order to determine the most appropriate combinations of observables from which to infer hydrometeor properties and behavior. The next decade should see the exploitation of the technique and related observations in many fields of study, including cloud physics, hydrometeorology, flash flood forecasting, severe storm nowcasting, microclimatology, air pollution scavenging, electromagnetic propagation and entomology.

REMOTE SENSING OF RANDOM MEDIA WITH ELLIPSOIDAL INHOMOGENEITIES

Ari Sihvola, Ismo V. Lindell

Helsinki University of Technology, Electromagnetics Laboratory, Otakaari 5 A, SF-02150 Espoo, FINLAND

ABSTRACT: In this paper, the polarizabilities of continuously inhomogeneous dielectric ellipsoids are analyzed. A differential equation is derived for the scattered potential, and from the amplitude of this function, the polarizability results. The limitation is that the permittivity function of the ellipsoid depend only on one coordinate (ξ) in the ellipsoidal coordinate system.

Keywords: Dielectric mixtures, continuous random medium, composite materials.

INTRODUCTION

The dielectric and magnetic properties of composite media are of importance in many applications where these materials are to be treated macroscopically homogeneous. The concept of effective permittivity is one notion that is much used in describing macroscopic properties of dielectric mixtures. It contains information about the average polarization in the composite material, being the ratio between the average displacement and the applied electric field. Averaging has to be made in scales smaller than the spatial variation of the incident electromagnetic field. This means that the wavelength of the applied field has to be larger than the characteristic distance of the inhomogeneities of the heterogeneous medium: this length is the size of the scatterers in case of the mixture being discrete (inclusions in a background medium), or the correlation length for the case of continuous random material. Therefore, effective permittivity is a low-frequency concept. At higher frequencies, scattering effects take over, and the term "effective permittivity" loses its meaning. However, the term "scatterer" is used in this paper denoting inclusions smaller than wavelength although scattering phenomena are neglected.

LAPLACE EQUATION IN THE ELLIPSOIDAL COORDINATE SYSTEM

Consider an inhomogeneous ellipsoid lying in a background medium of dielectric constant ϵ_0 and immersed in a static field. The incident static field is directed along the z -axis of the ellipsoid. This assumption does not entail loss of generality because an oblique field can be resolved into components along the axes and the resultant polarizability is a superposition of the three component polarizabilities.

In order to solve the internal and external fields of the problem, the solution of Laplace equation has to be known. The electric field is calculated as the negative gradient of a potential function ϕ :

$$\vec{E}(\vec{r}) = -\nabla\phi(\vec{r}) \quad (1)$$

In order for the Laplace equation to be separable in the geometry of the ellipsoid problem, the dielectric profile and the permittivity function may depend only on one coordinate in the ellipsoidal coordinate system: constant-permittivity surfaces have to be confocal (for ellipsoidal coordinates, see (Landau and Lifshitz, 1984) and (Kellogg, 1953)). These coordinates ξ, η, ζ are related to the Cartesian coordinates x, y, z by

$$\frac{x^2}{a^2+u} + \frac{y^2}{b^2+u} + \frac{z^2}{c^2+u} = 1 \quad (2)$$

being the three real roots of this cubic equation of u . The coordinate ξ is the root that lies in the range $\xi \geq -d^2$ where d is the smallest of the ellipsoid axes a, b , and c . Constant- ξ surfaces are ellipsoids all confocal to the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (3)$$

which corresponds to $\xi = 0$.

Let the permittivity function of the scatterer be only a function of the ξ coordinate:

$$\epsilon(\vec{r}) = \epsilon(\xi) \quad (4)$$

Let the incident uniform field be z -directed. The total field outside and inside the scatterer $-\nabla\phi$ is the sum of the incident field \vec{E} and the scattered field $-\nabla\phi_s$, where ϕ_s is the scattered potential, the amplitude of which is proportional to the dipole moment of the scatterer.

$$\nabla\phi_s(\vec{r}) = \vec{E} + \nabla\phi(\vec{r}) \quad (5)$$

The divergencelessness of the displacement,

$$\nabla \cdot \vec{D} = \nabla \cdot [-\epsilon(\xi)\nabla\phi(\vec{r})] = 0 \quad (6)$$

leads to the following equation

$$\nabla \cdot [\epsilon(\xi)\nabla\phi_s(\vec{r})] = \nabla\epsilon(\xi) \cdot \vec{E} = E\hat{z} \cdot \nabla\epsilon(\xi) \quad (7)$$

where \hat{z} is the unit vector in z direction. From the scattered potential, z has to be peeled off in order for the rest to be only a function of ξ :

$$\phi_s(\vec{r}) = x f_s(\xi) \quad (8)$$

Performing the differential operations in ellipsoidal coordinates, it can be seen after some algebra that the η and ζ dependencies cancel from the differential equation (7), the resulting equation for the ξ -dependent scattered potential being

$$2(\xi + a^2)\epsilon(\xi)f_s''(\xi) + \left[3 + \frac{\xi + a^2}{\xi + b^2} + \frac{\xi + a^2}{\xi + c^2}\right]\epsilon(\xi)f_s'(\xi) + 2(\xi + a^2)\epsilon'(\xi)f_s'(\xi) + \epsilon'(\xi)f_s(\xi) = \epsilon'(\xi)E \quad (9)$$

where a , b , and c are the axes of the ellipsoid in x , y , and z directions, respectively. This second-order ordinary differential equation for the unknown function $f_s(\xi)$ can also be written in the form

$$[\epsilon(\xi)(\xi + a^2)R(\xi)f_s'(\xi)]' + \epsilon'(\xi)\frac{R(\xi)}{2}f_s(\xi) = \epsilon'(\xi)\frac{R(\xi)}{2}E \quad (10)$$

where $R(\xi) = \sqrt{(\xi + a^2)(\xi + b^2)(\xi + c^2)}$. The surface of the scatterer is the $\xi = 0$ surface. ξ is negative inside the ellipsoid and positive outside it. On z axis, the connection between the coordinates is $z^2 = \xi + a^2$. The smallest value for ξ is $-a^2$ where a is the smallest of the three axes a , b , and c . The core of the ellipsoid, i.e. the surface where ξ has the minimum value, is an ellipse in the plane of the two major axes of the ellipsoid, and this ellipse shrinks to a point as the ellipsoid degenerates into sphere.

Given the permittivity profile $\epsilon(\xi)$, the solution can be found in a similar way as in the spherical case (Sihvola and Lindell, 1989). Solving the scattered potential inside the scatterer, for example, as a power series, leaves one undetermined constant. This constant can be enumerated by matching the inside solution to the outside solution. Outside the scatterer, where $\epsilon = \epsilon_0$, the solution is an integral (see equation (13) in the following). The amplitude of this function comes from the continuity of the potential across the boundary, and matching the derivatives gives the remaining undetermined constant. The derivative of the potential across the boundary is continuous if the permittivity of the scatterer decreases to the background value there. If the permittivity is not continuous across the boundary of the scatterer, the discontinuity of the potential can be calculated according to the continuity requirement of the normal displacement.

TESTS FOR THE DIFFERENTIAL EQUATION

The differential equation can be tested by investigating its solution in a homogeneous ellipsoidal region. There, $\epsilon(\xi)$ is constant and $\epsilon'(\xi) = 0$, and the equation (9):

$$f_s''(\xi) = -\left[\frac{3}{2(\xi + a^2)} + \frac{1}{2(\xi + b^2)} + \frac{1}{2(\xi + c^2)}\right]f_s'(\xi) \quad (11)$$

from which it is clear that the two solutions are, first, $f_s(\xi)$ is constant, and the second one is

$$f_s'(\xi) = -\frac{1}{(\xi + a^2)\sqrt{(\xi + a^2)(\xi + b^2)(\xi + c^2)}} \quad (12)$$

or

$$f_s(\xi) = \int_{\xi}^{\infty} \frac{ds}{(s + a^2)(\sqrt{(s + a^2)(s + b^2)(s + c^2)})} \quad (13)$$

which are correct (see (Landau and Lifshitz, 1984)).

The other test is to let the ellipsoid become a sphere, i.e. set $a = b = c$. Noting that $\xi + a^2 = \xi + b^2 = \xi + c^2 = r^2$, and that

$$\frac{\partial f_s}{\partial \xi} = \frac{1}{2r}f_s'(r) \quad (14)$$

$$\frac{\partial^2 f_s}{\partial \xi^2} = \frac{1}{4r^2}f_s''(r) - \frac{1}{4r^3}f_s'(r) \quad (15)$$

the equation (9) gives

$$r\epsilon(r)f_s''(r) + 4\epsilon(r)f_s'(r) + r\epsilon'(r)f_s'(r) + \epsilon'(r)f_s(r) = \epsilon'(r)E \quad (16)$$

which is correct because it is the same equation as the one coming from the scattered-potential equation in the radial geometry in spherical coordinate system:

$$\nabla \cdot [\epsilon(r)\nabla\phi_s(\vec{r})] = \nabla\epsilon(r) \cdot \vec{E} \quad (17)$$

with the separation $\phi_s(\vec{r}) = x f_s(r) = f_s(r)r \cos \theta$. From this, equation (16) follows immediately.

THE POLARIZABILITY

From the amplitude of the scattered field outside the ellipsoid, when $\xi > 0$, the polarizability in z direction, α^z can be calculated. The polarizability follows from the amplitudes of the scattered function and its connection to the dipole moment amplitude of the ellipsoid:

$$\alpha^z = \frac{\epsilon_0 V}{N^z E} f_s(0) \quad (18)$$

where N^z is the depolarization factor of the ellipsoid in z direction.

$$N^z = \frac{abc}{2} \int_0^{\infty} \frac{ds}{(s + a^2)\sqrt{(s + a^2)(s + b^2)(s + c^2)}} = \int_0^{\infty} \frac{ds}{(s + a^2)\sqrt{(s + a^2)(s + b^2)(s + c^2)}} \quad (19)$$

For values of polarization factors, see (Landau and Lifshitz, 1984), (Osborn, 1945), and (Stoner, 1945).

Knowing this polarizability, and the corresponding polarizability components in y and x directions, the anisotropic and isotropic effective permittivities of mixtures containing this type of scatterers (in aligned orientation and randomly oriented, respectively) can be calculated. How this is done, is shown in the next section.

THE EFFECTIVE PERMITTIVITY

The effective permittivity of a mixture containing inhomogeneous ellipsoids can be solved by quasistatic analysis according to (Sihvola and Kong, 1988). Let the background medium be of permittivity ϵ_0 as before, and let there be n ellipsoidal inclusions per unit volume. Consider first the case that all the ellipsoids are aligned equally, their corresponding z parallel. This leads to an anisotropic mixture, and the effective permittivity will be dyadic. In the following, consider fields and dipole moments in the z -direction.

Define the effective permittivity as the ratio between the average displacement \bar{D} and the incident electric field \bar{E} . On the other hand, the displacement can be calculated from the average polarization, and it is therefore

$$\bar{D} = \epsilon_{eff} \bar{E} = \epsilon_0 \bar{E} + \bar{P} \quad (20)$$

where the average polarization is equal to the dipole moment of one scatterer multiplied by their number density:

$$\bar{P} = n\bar{p} \quad (21)$$

The dipole moment of a scatterer is its polarizability times the exciting field,

$$\bar{p} = \alpha^* \bar{E}^* \quad (22)$$

The polarizability for ellipsoids in the z -direction is α^* . For the ellipsoid case, the polarizability depends on the direction. The exciting field is not the same as the macroscopic field but it depends on the permeating polarization itself and the form of the scatterer. Yaghjian (1980, 1985) has given the decomposition to calculate the exciting field as a function of the shape of the scatterer:

$$\bar{E}^* = \bar{E} + \frac{\bar{\bar{L}} \cdot \bar{P}}{\epsilon_0} \quad (23)$$

where $\bar{\bar{L}}$ is the unit-trace depolarization dyadic, or the source dyadic. The maximum chord length d_{max} of the scattering volume, that limits the validity of the decomposition (23) is given by $d_{max} \simeq \frac{\lambda}{2\pi}$ (see (Yaghjian, 1985)) where λ is the wavelength of the operating field. This imposes the quasistatic restriction for the size of the scatterers in the mixture.

For ellipsoidal scatterers with depolarization factors N^x, N^y, N^z , the depolarization dyadic is

$$\bar{\bar{L}} = N^x \hat{z}\hat{z} + N^y \hat{y}\hat{y} + N^z \hat{x}\hat{x} \quad (24)$$

where hat means the unit vector in the corresponding direction. The exciting field in the direction of z axis is

$$\bar{E}^* = \bar{E} + N^z \frac{\bar{P}}{\epsilon_0} \quad (25)$$

Hence, the effective permittivity in the z -direction is

$$\epsilon_{eff}^z = \epsilon_0 + \frac{n\alpha^z}{1 - N^z \frac{n\alpha^z}{\epsilon_0}} \quad (26)$$

and the dyadic effective permittivity is

$$\bar{\bar{\epsilon}}_{eff} = \epsilon_{eff}^x \hat{x}\hat{x} + \epsilon_{eff}^y \hat{y}\hat{y} + \epsilon_{eff}^z \hat{z}\hat{z} \quad (27)$$

If the layered ellipsoidal inclusions are randomly oriented in the mixture, there is no preferred direction and the effective permittivity is scalar. In this case the average polarization has to be calculated by averaging over the dipole moments in different directions, and the resulting effective permittivity is

$$\epsilon_{eff} = \epsilon_0 + \frac{\frac{1}{3} \sum_{i=x,y,z} n\alpha^i}{1 - \frac{1}{3} \sum_{i=x,y,z} N^i \frac{n\alpha^i}{\epsilon_0}} \quad (28)$$

REFERENCES

- L. D. Landau and E. M. Lifshitz: *Electrodynamics of continuous media*, Section 4, Second Edition, Pergamon Press, 1984.
- O. D. Kellogg: *Foundations of potential theory*, Chapter VII, Dover Publications, New York, 1953.
- J. A. Osborn: Demagnetizing factors of the general ellipsoid. *The Physical Review*, Vol. 67, No. 11-12, p. 351-357, 1945.
- A. H. Sihvola and J. A. Kong: Effective permittivity of dielectric mixtures. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 26, No. 4, p. 420-429, July 1988. See also: Corrections, Vol. 27, No. 1, p. 101-102, January 1989.
- A. Sihvola and I. V. Lindell: Polarizability and effective permittivity of layered and continuously inhomogeneous dielectric spheres. *Journal of Electromagnetic Waves and Applications*, Vol. 3, No. 1, p. 37-60, 1989.
- E. C. Stoner: The demagnetizing factors for ellipsoids. *Philosophical Magazine*, Ser. 7, Vol. 36, No. 263, p. 803-821, 1945.
- A. Yaghjian: Electric dyadic Green's function in the source region. *Proceedings of the IEEE*, Vol. 68, No. 2, p. 248-263, 1980.
- A. Yaghjian: Maxwellian and cavity electromagnetic fields within continuous sources. *American Journal of Physics*, Vol. 53, No. 9, p. 859-863, 1985.

USE OF CIRCULAR POLARIZATION IN A MARINE RADAR POSITIONING SYSTEM

D.G. Michelson, D.J. MacNeil, and E.V. Jull
 Department of Electrical Engineering,
 University of British Columbia, Vancouver, Canada V6T 1W5

R.G. Lyall and H.H. Lanziner
 Offshore Systems Ltd.,
 1974 Spicer Road, North Vancouver, Canada V7H 1A2

ABSTRACT

There is a need for a simple, reliable vessel positioning system for use by vessels navigating in coastal or inshore waters during periods of limited or restricted visibility. Existing aids to navigation either lack the accuracy required for use in narrow channels and harbours (e.g., LORAN-C or GPS) or are too expensive and require too much ongoing maintenance and support to be viable for general use (e.g., active microwave transponders). Here we describe a passive radar positioning system which uses a network of specially designed trihedral twist reflectors as landmarks and circular polarization for background clutter suppression.

Field trials of the system have been conducted in Indian Arm, British Columbia and Port aux Basques, Newfoundland. Initial results have confirmed that use of circular polarization can suppress background clutter sufficiently to permit positive identification of the specially designed radar targets by an automated system.

Key words: circular polarization, radar reflector, radar navigation, radar positioning.

INTRODUCTION

During the last ten years, the marine industry has expressed considerable interest in the development of automatic radar positioning systems for use by ships navigating in busy harbours or narrow channels during periods of inclement weather and/or reduced visibility. An automatic radar positioning system which employs passive reflectors has three principal advantages over conventional microwave transponder based positioning systems:

- * Shore-based passive reflectors have a low initial cost and should require minimal or no maintenance except in the unlikely event that a reflector sustains mechanical damage. Site services such as electrical power are not required;

- * Shore-based reflectors are not subject to failure due to loss of power or electronic malfunction; and,

- * Shipboard electronics do not have to be recalibrated when a shore-based reflector is repaired or replaced.

Passive reflector based radar positioning systems suffer from problems not normally

associated with microwave transponder based positioning systems, including:

- * Interference from ground and sea clutter returns;

- * Loss of signal due to multipath fading; and,

- * Limited angular response of targets.

In the late 1970's, the Canadian Dept. of Transport and the Dominion Marine Association evaluated a minicomputer based radar positioning system called PRANS [1]. In the early 1980's, researchers at McMaster University suggested several enhancements to the PRANS concept including the use of a novel trihedral twist reflector together with a simple depolarizing vector discriminant (transmission of horizontally polarized pulses and reception of both horizontally and vertically polarized returns) [2]. Since this system would require a radar with a dual-polarized antenna and a dual-channel receiver, industry observers expressed concern that the cost of the system would limit its acceptance. In this paper, we consider the use of circular polarization as an alternative target-clutter discriminant.

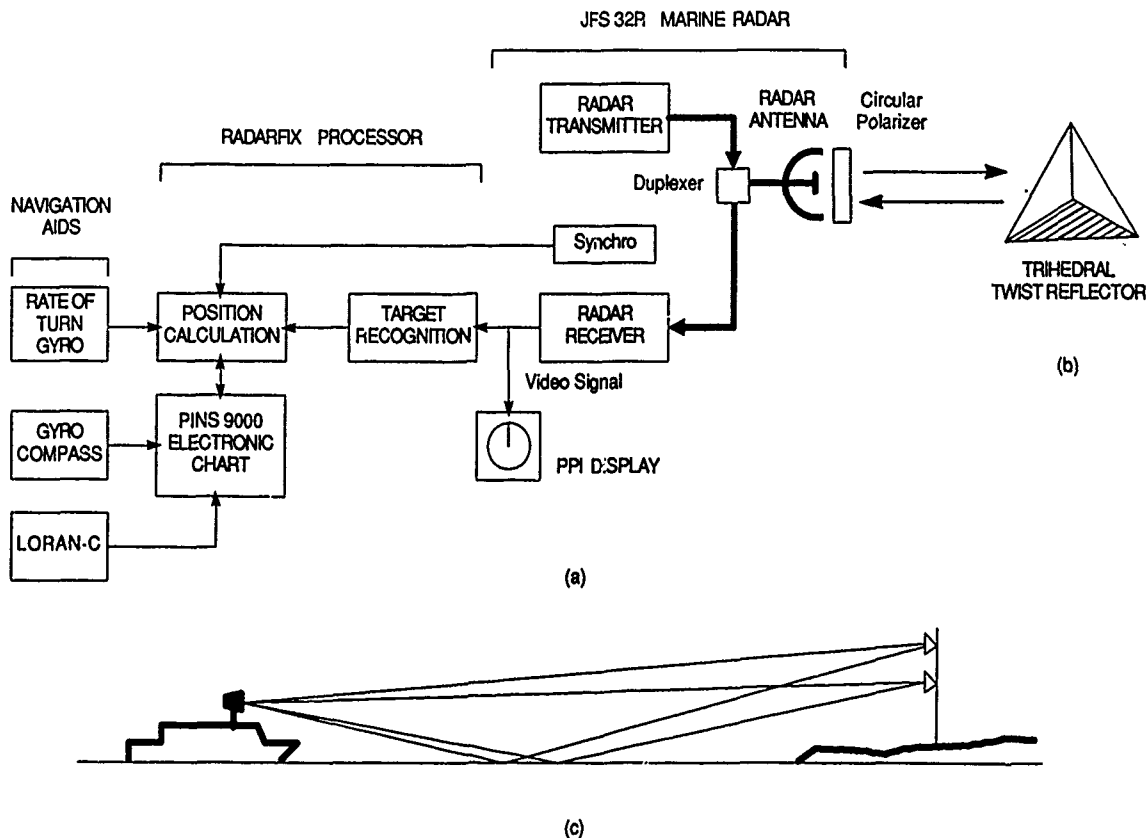


Fig. 1 - RANAV system overview. (a) Shipboard electronics. (b) Trihedral twist reflector. (c) Use of space diversity to reduce multipath fading.

SYSTEM DESCRIPTION

A prototype version of the RANAV system is shown in fig. 1(a). It consists of a JFS 32R 3 cm marine radar with a slotted waveguide array antenna and PPI display, a Radarfix processor [3] with a gyro rate-of-turn sensor attached, and an OSL PINS 9000 electronic chart with gyro compass and Loran-C receiver inputs. The horizontally polarized radar antenna was converted to circular polarization with the addition of a polarizing grid. The shore-based reflector shown in fig. 1(b) is a standard trihedral radar reflector with a twist reflector mounted along one face.

Although use of circular polarization normally reduces multipath fading substantially in transponder-based positioning systems, no such reduction is enjoyed by radar systems which employ circular polarization. Two reflectors can be mounted with a given vertical separation in a space diversity configuration (as suggested by fig. 1(c)) to reduce such fading. A simple theoretical model was used to optimize the vertical spacing given the location of the reflectors and their height above sea level. Although the reflectors have a fairly wide azimuthal beamwidth, it

is not always sufficient. In some cases, arrays of reflectors could be used to improve the angular coverage from a given reflector site.

TARGET-CLUTTER DISCRIMINATION ALGORITHMS

Many polarimetric target-clutter discrimination algorithms have been devised [4]. Circular polarization was considered for use in RANAV because it discriminates against odd-bounce reflectors (reportedly the major source of difficulty encountered in the PRANS trials) rather effectively while requiring relatively modest hardware. Use of more sophisticated target-clutter discrimination algorithms would require the use of dual polarized antennas and, in some cases, dual channel receivers. It is not yet clear that the increase in performance can be justified in light of the additional expense since an alternative would be to simply make the radar reflectors a little larger and thereby retain an acceptable signal to clutter ratio. Qualitative comparisons between circular and crossed linear polarization were made during the summer of 1988 in Vancouver harbour. Observers noted surprisingly little difference in clutter suppression between the two.

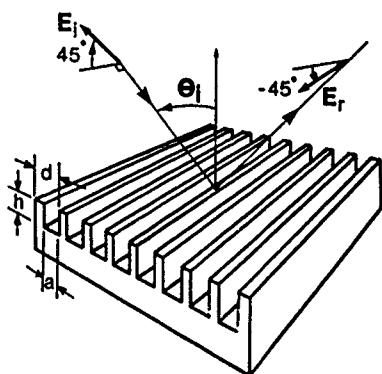


Fig. 2 - Rectangular groove twist reflector. Incident wave is linearly polarized at 45° to the grooves. The reflected wave is linearly polarized at -45° to the grooves for a net rotation of 90° .

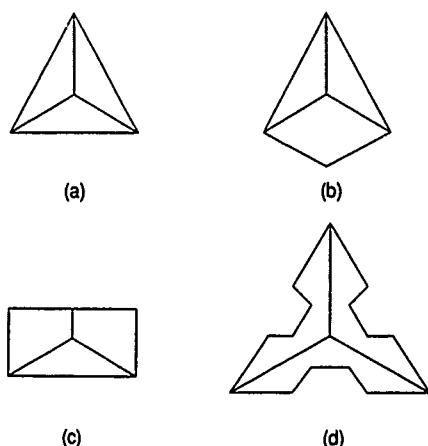


Fig. 3 - Alternative trihedral corner reflector geometries.

(a) Conventional triangular trihedral reflector. (b) Triangular trihedral reflector with a square base. (c) Truncated square trihedral reflector with a triangular base. (d) Compensated triangular trihedral reflector.

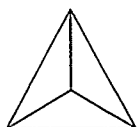
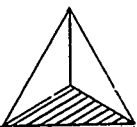
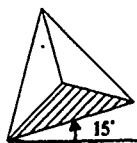
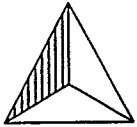
































TRIHEDRAL TWIST REFLECTOR DESIGN

Replacement of the wire grid twist reflector used in CRL's prototype with a rectangular groove twist reflector (fig. 2) has two advantages. First, the rectangular groove twist reflector is far less fragile than the wire grid and is potentially easier to manufacture. Second, the rectangular groove twist reflector has a wider operating bandwidth than the wire grid version which gives it a broader azimuthal response. A rigorous numerical analysis based on mode matching was used to determine the necessary design parameters [6].

The response pattern of a corner reflector can be altered by changing the shape of the reflector panels [7] as suggested by fig. 3. It is possible, for example, to broaden the azimuthal response of the reflector while sacrificing the elevation response pattern which is far less critical. An algorithm based on geometric optics was developed to predict the response pattern of the modified reflectors.

The polarization response of conventional and twist trihedral reflectors are compared for various orientations (i.e., various angles of rotation about their symmetry axes) in Table 1. Unlike conventional trihedral corner reflectors, trihedral twist reflectors preserve the sense of rotation of an incident circularly polarized wave and are therefore visible to circularly polarized radars. By suitably orienting the trihedral twist reflector, it can be made to either rotate linearly polarized waves by 90° or simply return them with their original polarization. In one application, navigation aids could be equipped with radar reflectors which would be visible to both horizontally and circularly polarized radars. Alternatively, shore targets used for positioning systems based on crossed linear polarization could also be used by systems which employ circular polarization.

Table 1 - Polarization response of conventional and twist trihedral reflectors for various orientations

|  | |  | |  | |  | |
|---|---|---|---|---|---|---|---|
| Incident Polarization | Reflected Polarization | Incident Polarization | Reflected Polarization | Incident Polarization | Reflected Polarization | Incident Polarization | Reflected Polarization |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

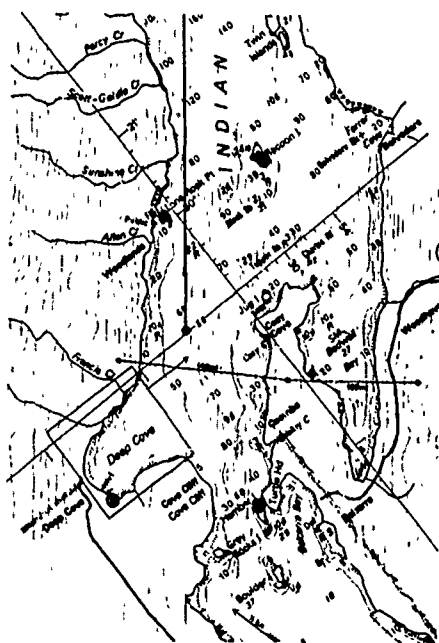


Fig. 4 - Chart of Deep Cove in Indian Arm, the site of the RANAV sea trials. The locations of three radar twist reflectors are shown: Raccoon Island, Deep Cove, and Hamber Island.

RESULTS

Sea trials to investigate the accuracy and reliability of RANAV were conducted in Deep Cove (a rugged, sparsely populated section of Indian Arm near Vancouver Harbour) in the summer of 1988 and spring of 1989. A network of three reflector sites was established as indicated in fig. 4. A conventional marine radar image of Deep Cove is shown in fig. 5(a). The radar image obtained when circular polarization is used and the radar receiver gain and video threshold are suitably adjusted is shown in fig. 5(b). The echoes from the three reflectors and the computer-generated flags which mark them are clearly visible.

The performance of RANAV was compared to a co-located microwave positioning system. RANAV was capable of repeatable accuracy to better than three metres within the area of reflector coverage. Static accuracy of less than one metre was typical.

In January 1989, RANAV began operational sea trials aboard Marine Atlantic's M.V. Atlantic Freighter, a 150 metre cargo ferry on the run between Newfoundland and Cape Breton. A network of RANAV reflectors has been installed at Port aux Basques, Newfoundland, the northern terminus of the run. Well known for its hazardous approach, high winds, and generally harsh environment, Port aux Basques will be a good test of RANAV's operational capabilities.



(a)



(b)

Fig. 5 - Radar images of Deep Cove. (a) Conventional marine radar image. (b) Radar image using circular polarization and suitable adjustment of receiver gain and video threshold. Radar returns from three reflectors are the only visible targets. The large blocks located next to each of the targets are computer generated flags.

REFERENCES

- [1] A.A. Hope, "Evaluation of a Precise Radar Navigation System (PRANS)", Transport Canada, TP 2980, Aug. 1981.
- [2] R. Cho, S. Haykin, and T. Greenlay, "Polarimetric Radar for Precise Navigation (PRAN)", McMaster University, Apr. 1986.
- [3] A.D. Virnot, "The Radarfix Positioning System", Canadian Hydrographic Conference, Burlington, Ontario, Feb. 1987.
- [4] J.L. Eaves and E.K. Reedy, ed., "Principles of Modern Radar", Van Nostrand Reinhold, 1987, pp. 619-645.
- [5] H. Mott, "Polarization in Antennas and Radar", John Wiley, 1986, pp. 239-243.
- [6] D.G. Michelson, P.Q.H. Phu, and E.V. Jull, "Millimetre-wave grating polarizers", 22nd General Assembly of URSI, Tel Aviv, Aug. 1987. p. 60.
- [7] S.D. Robertson, "Targets for microwave radar navigation", Bell Sys. Tech. J, vol. 26, no. 4, Oct. 1947, pp. 852-869.

We gratefully acknowledge the support provided by Transport Canada (Transportation Development Centre), Marine Atlantic, and the Science Council of British Columbia.

The Frequency and Polarization Dependence of Complex RCS Signatures

S. Riegger, D. Kähny, W. Wiesbeck*

Institut für Höchstfrequenztechnik und Elektronik
University of Karlsruhe, Kaiserstr. 12 D-7500 Karlsruhe, FRG

In previous papers several types of characteristic signatures, derived from coherent polarimetric Radar Cross Section (RCS) measurements were introduced [1]. These signatures are either 3 dimensional or, as cross sectional views of 3 dimensional ones, 2 dimensional. Up to now primarily first order signatures, which result from measured data by linear or geometrical operations, are used. Higher order signatures, e.g. depolarization or characteristic polarization states, are still difficult to compute or to interpret.

For the composition of 3-D signatures four variables are available as there are: polarization state (ellipticity, rotation) frequency and phase or amplitude. This means that for 3-D signatures there is always one variable not shown. For operational systems this parameter is most times the frequency. It is the intention of this paper to demonstrate the dependence of 3-D signatures from the fourth variable, especially the frequency. Coherent polarimetric measurements have been performed over the frequency range from 4 GHz to 20 GHz on several targets. The results of these measurements will be demonstrated in a film. The polarization signature (PS) [2] and the dispersive polarimetric signature (DPS) [1] will be shown.

The selected objects are :

- Dielectric cylinder
- Configurations of two cylinders
- Twig from a silver fir
- Rough surface with dielectric cylinders.

The objects were selected to increase the understanding of the information content of polarimetric signatures by starting with a very simple cylinder, proceeding with more complex targets to an object composed from a statistical rough surface with several dielectric cylinders on it. As an example the polarization signatures of the silver fir twig are shown in fig. 1 for several frequencies. In the film successively following signatures of the objects will be shown so that the impression of a steady growth versus the fourth variable is reached.

- [1] Kähny, D. et al.; "Coherent Polarimetric Signatures of Coniferous Trees, a Survey"; *Proc. 4. Colloq. on Spectral Sign. of Objects in Remote Sensing*; Aussois/France 1988
- [2] van Zyl, J. et al.; "Imaging Radar Polarization Signatures - Theory and Observation"; *Radio Science*; Vol. 22 1987

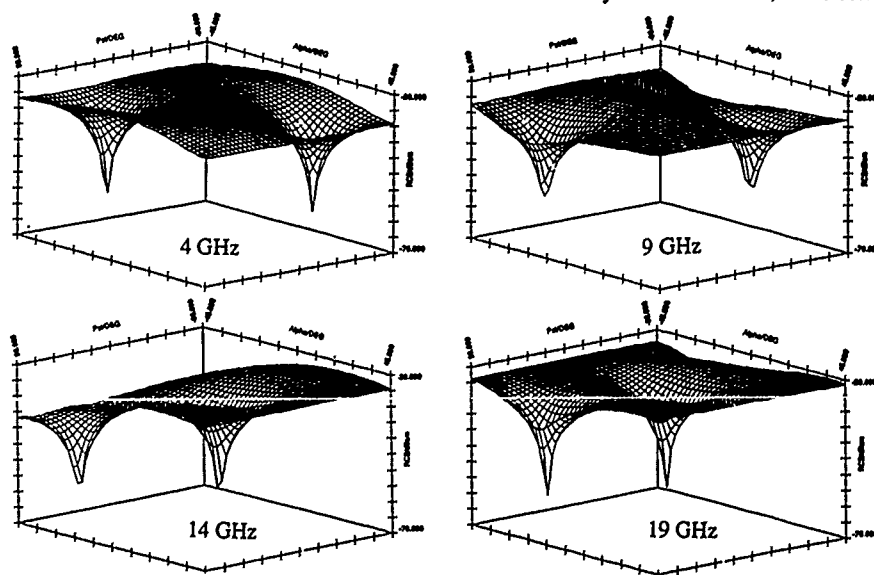


Fig. 1 Co-polarized Polarization Signatures of a silver fir twig at several frequencies

EFFECT OF TARGET SIZE AND TILT ON TARGET IDENTIFICATION BASED UPON FULL POLARIMETRIC RADAR IMAGING DATA

J. Richard Huynen

P. Q. Research
10531 Blandor Way
Los Altos Hills, California 94022, USA

ABSTRACT

Target analysis based on full polarimetric radar imaging data is still in a very rudimentary stage. The most popular methods go back some 30 years, when it was customary to present rectangular or circular plots of radar cross section (RCS) versus aspect angle, for horizontal (HH), vertical (VV) or cross polarized (HV) reception modes. If the target is unknown, usually very little concrete facts related to target reality is found. This is because the radar platform which defines the (HV) frame introduces an observation bias relative to the target's own angle of tilt, ψ , along the radar line of sight. Other observation biases may occur due to comparison of some type of targets which differ only in relative size. In this paper the following effects on real target returns are studied in detail:

- Effects of target tilt (ψ) and methods to eliminate target tilt-bias (desizing).
- Effects of target size and how it may be used in target sampling methods (desizing).
- Effects of target symmetries (physical and radar target symmetries) or the lack thereof.
- Effects of spurious target parameters typical for certain classes of targets (phyllotaxis, surface torsion, curvature and helicity).
- The consistent use of target decomposition theorems which are matched to a desired man-made or natural real-world situation and environment.
- The consistent use of power measures (Stokes vector, Stokes matrix, Wave and Target Coherency matrix) for both coherent and incoherent scattering behavior.
- The use of carefully chosen target significant parameters: A_0 , B_0 , B, C, D, E, F, G and H provides for a continuous range of values to designate targets. This is in contrast to, sphere-like or odd bounce, diplane-like or even bounce, older type of designations.

Key Words: Target scattering matrix, Mueller matrix, Wave and Target coherency matrix, polarimetric imaging; desizing and desizing; target reality.

INTRODUCTION

The objective of the present paper is to give a short review of new methods of data analysis, by using full polarimetric radar data, which aim at target identification. Most current efforts in this direction are still in a very primitive stage. Many view the objective of target identification as an operation on the incoming data stream, to be solved by computer software, with very little consideration or awareness about the physical significance the data may have, related to the reality of the target which is "out there." In order to improve concerns about physical target reality, one introduces models which are supposed to mimic real-world targets. Unfortunately, such models have limited use because of the variety of targets and different environments in which they are placed. One has to show how such models are relevant in each realistic case, and this is the problem one started out with. Hence, if model building leads to an infinite regress, what then could be the solution towards real-world target identification?

The approach the author offers here is model-free, nor does it assume a priori information about the target or its environment. Every world target is a potential object for investigation. The approach is based on simple properties of symmetry and non-symmetry which most targets must have and it emphasizes pre-processing of incoming data, so that so-called observation biases can be removed from the data base. Two such well-known biases are due to target tilt angle ψ and individual target size. Eliminating the effect of target tilt is called desizing, whereas removal of the effect of target size is called desizing of data.

THE TARGET SCATTERING MATRIX

In order to show how physical target tilt enters into the target scattering behavior we have to provide a short review already published elsewhere (Kennaugh, 1949-1954; Huynen, 1965, 1970, 1978, 1983). Let the target scattering matrix be given as:

$$S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} = \begin{bmatrix} a + b & c + id \\ c - id & a - b \end{bmatrix} \quad (1)$$

The SM is presumed given in the orthogonal (HV) horizontal-vertical reference frame. Because of reciprocity for the monostatic case, we can put $d = 0$. In practice one measures the SM by transmitting H and measuring S_{HH} , S_{HV} voltages in the two receiver channels,

then one switches to the vertical transmission mode and receives S_{VH} and S_{VV} . As soon as the SM is measured or is known, it becomes a pure target-property, quite independent of any type of wave illumination or reception!

THE TARGET ORIENTATION OR TILT PARAMETER

The target tilt ψ is defined by the solution of the following characteristic SM eigenvalue problem (Kennaugh, 1949-1954; Huynen, 1970, 1978, 1983)

$$S \hat{x} = s \hat{x} \quad (2)$$

The solution to the eigenvalue problem (2) defines a unitary matrix U , which is used to diagonalize the SM: Let

$$U = (\hat{x}_1, \hat{x}_2) \quad (3)$$

Since \hat{x}_1 and \hat{x}_2 form an orthogonal base, by which the SM is diagonalized, $S = U^* S_d U^{*T}$. The vector states \hat{x}_1 and \hat{x}_2 have physical significance. They define a unique "maximum-polarization" (or cross-pol null), $\hat{x}_1 = \hat{x}_m$ and \hat{x}_2 orthogonal to it. If \hat{x}_m is transmitted to the target (at fixed aspect angle at a given frequency, etc.), the target will return maximum power. The maximum power returned is a target characteristic. The characteristic wave \hat{x}_m is given in terms of physically meaningful geometric parameters; ellipticity τ_m and orientation ψ

$$\hat{x}_m = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \cos \tau_m \\ i \sin \tau_m \end{bmatrix} \quad (4)$$

If the target is rotated about the line of sight, the target tilt will change and hence with it ψ . Thus we have found an intrinsic measure of target tilt independent of a priori knowledge. Once we know how the tilt is formulated in the SM it is possible to eliminate it from the data base!

TARGET SYMMETRY AND NON-SYMMETRY

A symmetrical radar target is defined as an object which, when observed by radar, has a plane of mirror symmetry going through the line of sight direction. Hence any roll-symmetrical physical object also is radar target symmetric at all aspect angles and tilt angles (and at all frequencies). However there are objects which have mirror-symmetry only at certain prescribed aspect angles. Hence a trihedral has many physical planes of mirror-symmetry but only when that plane contains the radar line of sight observation axis, is that object radar target symmetric. We can derive an important property for radar symmetric targets: $\tau_m = 0$. Hence the maximum polarization (or cross-pol null) must be linear for symmetric radar targets. The SM for radar symmetric targets (i.e., at a selected observation angle) is when $\psi = 0$:

$$S S = \begin{bmatrix} a + b & 0 \\ 0 & a - b \end{bmatrix} \quad (5)$$

Here a belongs to the unit matrix in (5), which is the case of specular return, i.e., a sphere, any convex surface or odd-bounce gives this type of return. Below we will call $2A_0 = |a|^2$ the generator of target symmetry, which is a much wider concept (defined by $\tau_m = 0$). Conversely, $\tau_m = \pm 45^\circ$ determines a target where $a = 0$. Such a target is highly non-symmetrical and is called an N-target. N-targets play an important role in the target decomposition theorem (see below). Hence the scattering matrix for an N-target is:

$$S^N = \begin{bmatrix} b & c \\ c & -b \end{bmatrix} \quad (6)$$

Below we will show that the generator for the N-target is $B_0 - B = |c|^2$. The term $|b|^2$ in the SM (1) plays an intermediary role, it contributes both to target symmetry as well as to target-nonsymmetry.

POWER-RELATED TARGET PARAMETERS

The target scattering matrix has obvious importance because it represents the target-input data. It represents the field-description of the target in amplitude and phases. However, for the use of target data as target-identification parameters, a more suitable representation of data exists in terms of power. In the first case power eliminates the absolute phase for the target, i.e., the targets become "disconnected" because of independence of local position changes relative to the radar site location. Hence power-related parameters are incoherently additive. This property is very important because the concept of object is independent of local placement in the environment. It turns out that by computing power P from received voltage V (i.e., $P = |V|^2$) the auto- and cross-products: $|a|^2$, $|b|^2$, $|c|^2$, a^* , a , b^* and b are formed. It is clear that the product terms (when averaged) produce information about auto- and cross-correlations between SM-elements. Hence the average power $\langle P \rangle$ produces completely different kinds of target information. For example, we know that $|a|^2$ in the SM produces the specular (odd-bounce) term, but we observe below that in the power presentation $|a|^2$ becomes a generator of five real parameters, given by $|a|^2$, a^* and a . In fact, by this method $|a|^2$ is recognized not only as specularly but as the generator of radar target-symmetry. Conversely the parameter $|c|^2$ (after elimination of ψ) becomes the generator of target non-symmetry. A very important new insight into target structure is thus created:

A GENERAL RADAR TARGET MAY BE VIEWED AS CONSISTING OF A PART WHICH IS SYMMETRIC AND A PART WHICH IS NON-SYMMETRIC, WITH COUPLING TERMS BETWEEN THEM.

These concepts are derived in detail next: we introduce the Target Stokes matrix (Mueller-matrix)

$$M_0 = \begin{bmatrix} A_0 + B_0 & F & C & H = 0 \\ F & -A_0 + B_0 & G & D \\ C & G & A_0 + B & E \\ H = 0 & D & E & A_0 - B \end{bmatrix} \quad (7)$$

The coordinate-frame in which M_0 is given is such that

$$P = M_\psi g(a) \cdot h(b) \quad (8)$$

gives the back scattering power, if $g(a)$ is the Stokes vector for transmit polarization:

$$g(a) = \begin{bmatrix} p^2 \\ p^2 \sin 2\tau_a \\ p^2 \cos 2\tau_a \cos 2\psi_a \\ p^2 \cos 2\tau_a \sin 2\psi_a \end{bmatrix} = \begin{bmatrix} 8_0 \\ 8_1 \\ 8_2 \\ 8_3 \end{bmatrix} \quad (9)$$

and $h(b)$ has the same form for the receiver polarization. The Stokes matrix M_0 differs from M_ψ by the fact that $\psi = 0$, i.e., the target tilt angle has been eliminated, the target has been rotated such that its "axis" is aligned with the (HV) radar platform. For that case we

find in (7) the important rule: $H = 0$. There is a very simple relationship between the carefully chosen nine target parameter designations in (7) and the SM coefficients:

$$\left. \begin{aligned} |a|^2 &= 2A_0, & a b^* &= C - i D \\ |b|^2 &= B_0 + B, & b c^* &= E + i F \\ |c|^2 &= B_0 - B, & a c^* &= H + i G \end{aligned} \right\} \quad (10)$$

Equations (10) apply in general, with ψ or with $\psi = 0$. These parameters carry basic and general significance related to target reality. Equations (10) suggest that an alternative representation for the Target Stokes Scattering Operator must exist in the form of a 3×3 Target Coherency matrix (Huynen, 1985):

$$(a, b, c)^T (a, b, c)^* = \begin{bmatrix} |a|^2 & a b^* & a c^* \\ b a^* & |b|^2 & b c^* \\ c a^* & c b^* & |c|^2 \end{bmatrix} \quad (11)$$

There is a one-to-one correspondence between the complex matrix (11) and the real Stokes matrix (7).

We notice that (11) is hermitian and the target generators $|a|^2$ and $|c|^2$ for symmetry and non-symmetry are prominently displayed as diagonal terms. The third term $|b|^2$ plays an intermediate role and is called "the generator of target irregularity." Thus far we focused on target generators. But the off-diagonal terms are equally important as real target designators.

Parameters H and G are "COUPLING" terms. Parameters C and D are measures of **TARGET SHAPE**, C is global shape, i.e., $C = 0$ for a sphere, but C is large for a line element or a wire, D is related to "local shape," it has been identified with local curvature difference ($K_x - K_y$) on a convex surface patch for high-frequency scatter (Bennett et al., 1973; Boerner et al., 1981; Foo, 1982; Mieras, 1984; Chaudhuri et al., 1986). The parameter E is related to surface torsion (Huynen, 1988a); it is part of target non-symmetry. F is easily identified as **HELICITY**, being the difference between (LC, LC) and (RC, RC) circular polarized returns. Hence the full scope of target parameters is given by the following table:

TABLE OF BASIC TARGET PARAMETERS

| | |
|-----------|--|
| A_0 | = Generator of Target Symmetry |
| B_0 | = Generator of Target Structure (Hamiltonian) |
| $B_0 - B$ | = Generator of Target Non-Symmetry |
| $B_0 + B$ | = Generator of Target Irregularity |
| H_ψ | = Coupling due to Target Tilt Angle ψ |
| G | = Coupling between Symmetric/Non-Symmetric Parts |
| C | = Shape Factor (Maximum for Line Target) |
| D | = Measure of Local Curvature Difference |
| E | = Surface Torsion |
| F | = Target Helicity |

TARGET DECOMPOSITION THEOREM

In this report several methods for polarimetric target averaging are investigated. For most target identification problems one look at a composite target does not give enough information. Instead, for an ensemble of targets, sample Stokes matrix averages $\langle M \rangle$ are obtained. We know from elementary theory (Kennaugh, 1949-1954) that an SM for a single look target is given by only five independent parameters,

but the general averaged Stokes matrix has nine independent target parameters associated with it. Hence it is in general impossible to assign a SM to the average target mixture $\langle M \rangle$.

What remains central in the knowledge of target mixtures is the idea of an "effective single average object," which does have a unique SM equivalent. Hence this author has advocated a second approach which is that of target decomposition (Huynen, 1970, 1978,):

$$\begin{aligned} \langle M \rangle &= M(\text{ave}) + N(\text{residue}) \\ 9 \text{ par.} & \quad 5 \text{ par.} \quad 4 \text{ par.} \end{aligned} \quad (12)$$

The procedure for target decomposition is as follows: First one starts with $\langle M \rangle$ data given by $A_0, B_0, B, C, D, E, F, G$ and H . From these one computes average target (T) parameters: $A_0^T, B_0^T, B^T, C^T, D^T, E^T, F^T, G^T$ and H^T , where $A_0^T = A_0, C^T = C, D^T = D, G^T = G$ and $H^T = H$ (Huynen, 1970, page 166). Hence the N-target parameters are formed from $A_0^N = 0, B_0^N = B_0 - B_0^T, B^N = B - B^T, C^N = 0, D^N = 0, E^N = E - E^T, F^N = F - F^T, G^N = 0$ and $H^N = 0$.

The N-target residue is chosen such that it represents purely non-symmetrical target parameters. The originally nine independent parameters on the left have thus been split into a set composed of the average single target (with corresponding SM) given by five independent parameters and four parameters for the N-target.

The N-target has the desirable property that as a class of targets, it does not change with target tilt ψ and from this it follows that the process of target decomposition (12) itself does not change with target tilt (rotation of the radar reference frame about the line of sight direction). This method of target decomposition will be tested in this paper for different averaging procedures. These new averaging processes determine the effects of individual target tilts and of individual target sizes on the sampled matrix $\langle M \rangle$. Hence one can study the unique N-target-decompositions for various classes of targets and averaging procedures.

THE PROCESSES OF DESYING AND DESIZING OF INITIAL DATA

From the measured SM data, the Stokes-matrix parameters for single-look targets are computed from Eqs. (10). From this are found the target parameters: $A_0, B_0, B, C, D, E, F, G$, and H_ψ . Without removal of the ψ dependence, the parameters have little physical sense, except A_0, B_0 and F which are independent of target tilt angle ψ . The desying process is very simple, and is found in Huynen (1970). We thus find the desyed parameters: $A_0, B_0, B, C, D, E, F, G, H = 0$ and ψ . The method above is called desying of target parameters or the process $H = 0$.

Another averaging process was studied, which is based on the idea that individual targets can be different not only in tilt angle ψ , but also in size. Comparing two objects which have an identical shape, but which are different in size alone, leads one to look for a method of desizing individual targets. Obviously the difference in size cannot exceed the limits posed by the EM-wavelength, by which the objects are being observed.

How to desize an individual target return is not as obvious as it may look. One demands a parameter for overall target size by which the data may be normalized. However several candidate parameters for doing so

present themselves. Here some sophistication about understanding of the target return process might be helpful. An obvious device would be to take the span for such a measure and indeed some authors would favor such a choice (Boerner, 1981; Van Zyl et al., 1988). However, there are numerous other choices possible, each with a claim to fame to represent target size. If $\text{Span} = A_0 + B_0$, then other possibilities are either A_0 or B_0 separately. Obviously also a combination of desizing and desizing is possible!

EXPERIMENTAL RESULTS

The various averaging schemes for desizing and desizing or mixtures of these were tested on a sample, provided by Prof. F.T. Ulaby from the University of Michigan, of 26 single target SM data for tree canopy alone and trees with trihedral inserted (Ulaby, 1988). For measures of size were chosen the span: $A_0 + B_0$, also $2A_0 + B_0$ and B_0 alone, combined with desizing or not desizing.

A surprising result was found (Huynen, 1988b) that desizing together with B_0 as measure of target size provided the clearest decomposition into effective target and N-target residue. In fact it was found that the N-target residue for both trees alone and trees + trihedral, under these conditions, were almost identical. This clearly puts the N-target residue as part of the tree-canopy clutter alone, unaffected by the insertion of the trihedral.

CONCLUSION

Sixteen samples of single target full polarimetric SM data were analyzed using Stokes matrix averaging techniques. One set consisted of measurements of a tree canopy while the second set of samples were for trees with a trihedral inserted in the foliage. The sampling procedure covered four cases with desizing (removal of tilt angle ψ) and desizing (removal of effect of target size on the input data). The mixed Stokes matrix was then subjected to the target decomposition theorem (Huynen, 1970, Chapter 7), which separates from the target mixture an average single target + N-target residue.

Of the four methods of averaging, the combined effect of desizing and desizing shows most clearly the "symmetrizing" of the effective average target due to insertion of the trihedral. The N-target residue is shown to be largely part of the clutter background alone. This type of analysis clearly shows the benefit of using parameters related to target reality, over conventional methods. Here a list of parameters with a continuous range of numerical values is given by which to investigate real-world target structure.

ACKNOWLEDGEMENTS

The author is grateful to Prof. F.T. Ulaby from the University of Michigan, for generously supplying him with measured X-band SM data on tree canopies and trees imbedded with a trihedral.

REFERENCES

- Bennett, C.L., Auckenthaler, A.M., Smith, R.S. and Delorenzo, J.D., "Space-time integral equation approach to the large body scattering problem", Sperry Research Center, Sudbury, MA, Final Report on Contract No. F30602-71-C-0162, AD763794, May 1973.
- Boerner, W.M., "Use of polarization in electromagnetic inverse scattering", Radio Science, Vol. 16, No. 6, pp1037-1045, Nov./Dec. 1981.
- Boerner, W.-M., Ho, C.-M. and Foo, B.Y., "Use of Radon's projection theory in electromagnetic inverse scattering", IEEE Trans. Antennas Propagation (Special Issue on Inverse Methods in Electromagnetics), Vol. AP-29, March 1981.
- Chaudhuri, S.K., Foo, B.Y. and Boerner, W.-M., "A validation analysis of Huynen's target-descriptor interpretations of the Mueller matrix elements in polarimetric radar returns using Kennaugh's physical optics impulse response formulation", IEEE Trans. on Ant. and Prop., pp1-20, Jan. 1986.
- Foo, B.Y., "A high frequency inverse scattering model to recover the specular point curvature from polarimetric scattering data", M.Sc. Thesis, Electr. Engr. & Comp. Sci. Dept., University of Illinois at Chicago, IL, Communications Lab Report No. 82-05-21, May 21, 1982.
- Guili, D., "Polarization diversity in radars", Proc. of IEEE, Vol. 74, No. 2, pp245-269, Feb. 1986.
- Holm, W.A., "Polarimetric fundamentals and techniques", in Principles of Modern Radar, edited by J.L. Eaves and E.K. Reedy, pp621-645, Van Nostrand Reinhold Co., 1987.
- Huynen, J.R., "Measurement of the target scattering matrix", IEEE Proc., Vol. 53, pp936-946, Aug. 1965.
- Huynen, J.R., "Phenomenological theory of radar targets", Doctoral Thesis, Technical University, Delft, The Netherlands, 1970. (Obtainable from the author, revised edition Nov. 1987.)
- Huynen, J.R., "Phenomenological theory of radar targets", Chapter 11 in Electromagnetic Scattering, edited by P.L.E. Uslenghi, Academic Press, New York, 1978.
- Huynen, J.R., "Towards a theory of perception for radar targets", in Inverse Methods in Electromagnetic Imaging, edited by W.-M. Boerner, pp797-822, D. Reidel Publishing Co., Dordrecht, The Netherlands, 1985.
- Huynen, J.R., "The calculation and measurement of surface torsion by radar", P.Q. Research, Report No. 102, June 1988a.
- Huynen, J.R., "Extraction of target-significant parameters from polarimetric data", Report No. 103, P. Q. Research, July 1988b.
- Kennaugh, E.M., "Effects of type of polarization on echo characteristics", O.S.U. Antenna Lab., Columbus, OH, Reports 389-1 to 389-24, 1949-1954.
- Mieras, H., "Local influence technique in time domain scattering", Ph.D. Dissertation, Northeastern University, Boston, MA, Feb. 1984.
- Ulaby, F.T., Moore, R.K. and Fung, A.K., Microwave Remote Sensing, Vol. III, Artech House Inc., 1986.
- Ulaby, F.T., Backscatter from a Tree Canopy, Data and Private Communications, Letter, January 1988.
- Van Zyl, J.J., Papas, C.H. and Elachi, C., "On the optimum polarizations of incoherently reflected waves", IEEE Trans. on An. and Propag., Vol. AP-35, No. 7, pp818-825, July 1987.

OPTIMIZATION PROCEDURES FOR SCATTERING MATRICES IN THE COHERENT
AND PARTIALLY COHERENT CASES

Wolfgang-M. Boerner, Wei-Ling Yan and Alexander B. Kostinski

University of Illinois at Chicago
UIC-EECS/CL, M/C 154
840 W. Taylor St., SEL-4210
Chicago, IL 60680-4348
Tel: +1(312)996-5480
Fax: +1(312)413-0024

Basic aspects of polarimetric radar/lidar target scattering for both the coherent and partially coherent cases are re-examined. The optimization procedures for the coherent case is formulated in terms of a "Three-Stage Procedure" for the general bi-static case. The distinction between the physics of partially polarized versus coherent waves is emphasized leading to a unique formulation for optimal reception of partially polarized waves scattered off a fluctuating ensemble of scatterers of known (measured) Mueller matrices. Expressions for total available intensity versus adjustable polarization-dependent (coherent) intensity are derived using the coherency matrix approach. By formulating a proper covariance matrix, it is shown how the measured Mueller matrix can be tested on measurement errors and how the partially polarized case reduces to the coherent case resulting in a unique set of optimal polarizations.

RECENT ADVANCES IN AIRBORNE TERRESTRIAL REMOTE SENSING WITH THE NASA
AIRBORNE VISIBLE/INFRARED IMAGING SPECTROMETER (AVIRIS),
AIRBORNE SYNTHETIC APERTURE RADAR (SAR), AND
THERMAL INFRARED MULTISPECTRAL SCANNER (TIMS)

Gregg Vane, Diane L. Evans and Anne B. Kahle

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ABSTRACT

Significant progress has been made in terrestrial remote sensing from the air with three NASA-developed sensors that collectively cover the solar-reflected, thermal infrared and microwave regions of the electromagnetic spectrum. These sensors are the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Vane, 1987), the Thermal Infrared Mapping Spectrometer (TIMS) (Kahle et al., 1980), and the Airborne Synthetic Aperture Radar (SAR) (Held et al., 1988), respectively. AVIRIS and SAR underwent extensive in-flight engineering testing in 1987 and 1988 and are scheduled to become operational in 1989. TIMS has been in operation for several years. In this brief paper, the sensors are described. At the symposium, results will be presented from recent experiments in the earth sciences conducted with AVIRIS, SAR and TIMS.

THE SENSORS

AVIRIS is a whiskbroom imaging sensor that is flown on the NASA ER-2 research aircraft. From an altitude of 20 km, AVIRIS covers a swath on the ground of 10.5 km at a ground instantaneous field of view of 20 m. Including a cross- and along-track spatial oversampling of 17 percent, the AVIRIS image spans 614 pixels. Two hundred and ten inherently registered images are acquired simultaneously in 10-nm-wide contiguous spectral bands spanning the entire solar reflected portion of the spectrum from 0.4 to 2.5 μm . This is accomplished by transmitting light from the scene collected by a common scanning foreoptic to 4 spectrometers, each with a line array of detectors in its focal plane. Data are encoded at 10 bits and recorded on-board the aircraft on a high density tape. The raw data rate of the sensor is 17 Mbps. Signal-to-noise performance referenced to a standard radiance model with a surface reflectance of 0.5 viewed at mid-latitude in mid-summer

through an atmosphere with rural aerosols and 23 km visibility is: for the visible from about 500 to 700 nm, 200:1; for the near infrared from about 700 to 1200 nm, from 100 to about 150:1; for the short wavelength infrared from 1200 to 1800 nm, 100:1, and from 1800 to 2500 nm, about 30:1. AVIRIS will be flown extensively in 1989 in support of research in atmospheric science, botany, geology, hydrology and oceanography.

TIMS is also a whiskbroom imager with a scanning foreoptic and set of 6 discrete detectors, each with an interference filter. The sensor covers the thermal infrared from 8 to 12 μm in nearly contiguous spectral bands centered approximately at 8.3, 8.7, 9.1, 9.8, 10.5, and 11.4 μm . The high sensitivity of the sensor results in a noise equivalent change in temperature of 0.1 to 0.3 degrees C. TIMS has an instantaneous field of view of 2.5 mrad and a field of view of 80 degrees. It is flown on a NASA Learjet and C-130 over a wide range of altitudes, and has been in operation for six years, supporting research in geology and botany. Additional experiments are planned in 1989 in atmospheric characterization.

The Airborne SAR is a 3-frequency, 4 polarization synthetic aperture radar that is flown on the NASA DC-8. The sensor images in C-band (5300 MHz), L-band (1225 MHz), and P-band (440 MHz) simultaneously in HH, HV, VH and VV polarizations. The system has a single STALO clock and a single exciter source, which operates at L-band. Each generated chirp is converted up to C-band and down to P-band for amplification by a single TWT for the P- and L-band frequencies, and 2 separate TWTs at C-band before being split into H and V channels for transmission. The H and V transmission events are separated in time by half a pulse repetition interval. In this way, the chirp waveform for each frequency/polarization is kept constant, since the circuitry is held in common as much as possible. The high resolution imagery is composed of 4000 pixels in azimuth by 750 pixels in range, with nominal 4 m and 10 m resolutions in

azimuth and slant range respectively. The raw data rate is 80 Mbps. SAR has been used in support of experiments in sea ice classification, geology, botany, ocean wave research, and calibration. It will be flown to Europe in the summer of 1989 in support of experiments in all these areas.

ACKNOWLEDGEMENTS

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

REFERENCES

- Held, D.N., Brown, W.E., Freeman, A., Klein, J.D., Zebker, H., Sato, T., Miller, T., Nguyen, Q., and Lou, Y., "The NASA/JPL multifrequency, multipolarization Airborne SAR system," in Proceedings of the IGARSS'88 Symposium, ESA SP-284 (IEEE 88CH2497-6), pp 345-349, 1988.
- Kahle, A.B., Madura, D.P., and Soha, J.M., "Middle infrared multispectral aircraft scanner data: Analysis for geological applications," Applied Optics, pp 2279-2290, 1980.
- Vane, G., ed., "Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): A description of the sensor, ground data processing facility, laboratory calibration, and first results," JPL Publication 87-38, 97 pp, 1987.

MINERAL IDENTIFICATION BY THE AVIRIS DATA

I. PIPPI

C.N.R. - I.R.O.E.
Florence - ITALY

ABSTRACT

The geological investigation is one of the most interesting application fields of the remote sensing techniques.

The new airborne imaging spectrometers give additional possibilities, allowing the mineral identification.

One of these sensors was developed at JPL, and was flown on board of the NASA U-2 aircraft, during the first measurement campaign in the summer 1987, over a site containing both hydrothermally altered and unaltered rocks well exposed at the surface.

Using the software developed at our Institute, any image of the site in any of the 210 available spectral bands or the spectral signature of any pixel can be displayed. The spectral data carry information corresponding to the composition of the ground being viewed and the intervening atmosphere.

Particular attention is paid to evaluate the radiometric calibration accuracy and the atmospheric effects utilizing the LOWTRAN 6 computer code.

First results of mineral identification are presented and discussed.

Keywords: AVIRIS, Imaging spectrometers, Mineralogy.

1. INTRODUCTION

The geological investigations by means of airborne or spaceborne sensors started fifteen years ago with the launch of the first Landsat. The limonite, a combination of Fe^{3+} minerals, was the first and only rock material identified from the Multispectral Scanner System data analysis and this capability was used for mapping hydrothermally altered rocks.

The addition of the $1.65\mu m$ and $2.22\mu m$ bands on the Thematic Mapper allowed the detection of the classes of minerals that have OH, HOH, or CO_3 absorption features, but it was impossible to discriminate the minerals inside the classes. Only the use of high spectral resolution remote sensing can allow the mineral identification.

For this purpose during the last decade same spectrometers have been flying on board of aircrafts or spacecrafts. The most promising seems to be the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) (Ref.1), an airborne prototype for the High-Resolution Imaging Spectrometer (HIRIS) (Ref.2) planned for launch in the mid 1990s on the polar-orbiting space platform.

The AVIRIS operates in the whisk-broom imaging mode with a spectral coverage from 0.40 to $2.45\mu m$ divided into 210 adjacent spectral bands.

In the first season of operation during the summer 1987, the AVIRIS flew on board of the NASA U-2 aircraft acquiring data over more than 30 different sites located in the United States. A site was located in the Cuprite mining district in western Nevada.

2. SITE DESCRIPTION

The Cuprite area is well known from a geological point of view and contains both hydrothermally altered and unaltered rocks, well exposed and nearly devoid of vegetation (Ref.3).

In particular the eastern half of the district is an area of extensive hydrothermal alteration within a sequence of rhyolitic welded ash flow and air fall tuffs. The altered units consist of a central core of almost pure silica, a ring of opalized rocks containing alunite and kaolinite, and an outer argillized zone containing mainly kaolinite, montmorillonite, and opal, and some limonite.

3. DATA EVALUATION

The radiometrically corrected image set, acquired over the Cuprite district, is obtained from the Jet Propulsion Laboratory, allowing us to evaluate the ability of the AVIRIS for the identification of several minerals.

The image set forms a data cube of which two axes represent spatial dimensions and the third represents the spectral dimension. The size of the data cube is 512 pixels along the flight direction by 614 pixels along the cross-track direction, with a ground instantaneous field of view of 20 meters, by 210 spectral bands, ranging from 0.40 to 2.45 μm with a spectral resolution of 9.8 nm. The data, in the format of 16 bit words, have a resolution of 10 bits.

Using the software developed at our Institute, any image of the site in any of the 210 available spectral bands or the spectral signature of any pixel can be displayed. In particular four images, each from a different AVIRIS spectrometer, are considered to evaluate the data quality.

Several periodic noises are present in the image data in addition to the random noise. These types of noise are already detected taking into account the images themselves. The mechanical vibrations and electronics interferences seem to be responsible for the rise of the pattern noise.

Another way to evaluate the data quality is to consider the image histograms. An example is given in Figure 1: (A) for the spectral band n.30, centered at 0.684 μm ; (B) for the spectral band n.68, centered at 1.057 μm ; (C) for the spectral band n.125, centered at 1.615 μm ; (D) for the spectral band n.188, centered at 2.233 μm .

If the spectral data related to a certain pixel are considered, they carry information corresponding to the composition of the ground being viewed and the intervening atmosphere, as shown in Figure 2 for the overall AVIRIS spectrum and in Figure 3 for the spectrum between 1.9 and 2.5 μm .

Unfortunately the signal-to-noise ratio of the available data decrease too rapidly from the visible spectral bands to the infrared ones, reaching a value around 20:1 at 2.2 μm (R:f.4).

In order to evaluate the atmospheric contribution, the LOWTRAN 6 computer program is used (Ref.5). The transmittance of the atmosphere, computed in the spectral region and in according to the parameters of our interest listed in Table 1, is reported in Figure 4.

The comparison between graphs as in Figure 3 and 4 allows first mineral identifications.

4. CONCLUSIONS

The imaging spectrometry seems to offer the possibility of identifying the minerals containing iron as well as those bearing OH, CO₃, and SO₄.

Nevertheless the improvement in signal-to-noise performances of the AVIRIS, specially in the infrared bands, should allow mapping of alteration area.

5. REFERENCES

1. "Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) - A Description of the Sensor, Ground Data Processing Facility, Laboratory Calibration, and First Results", G.Vane editor, JPL Publication 87-38, 1987.
2. "HIRIS High-Resolution Imaging Spectrometer: Science Opportunities for the 1990s", Instrument Panel Report, Earth Observing System Vol.IIc, NASA, 1987.
3. M.J.Abrams, R.P.Ashley, L.C.Rowan, A.F.H.Goetz and A.B.Kahle, "Mapping of hydrothermal alteration in the Cuprite mining district, Nevada, using aircraft scanner images for the spectral region 0.46 to 2.36 μm ", *Geology*, 5, pp.713-718, 1977.
4. "Proceedings of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Performance Evaluation Workshop", G.Vane editor, JPL Publication 88-38, 1988.
5. F.X.Kneizys, E.P.Shettle, W.O.Gallery, J.H.Chetwynd, L.W.Abreu, J.E.A.Selby, S.A.Clough and R.W.Fenn, "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 6", AFGL-TR-83-0187 Environmental Research Papers, n.846, 1983.

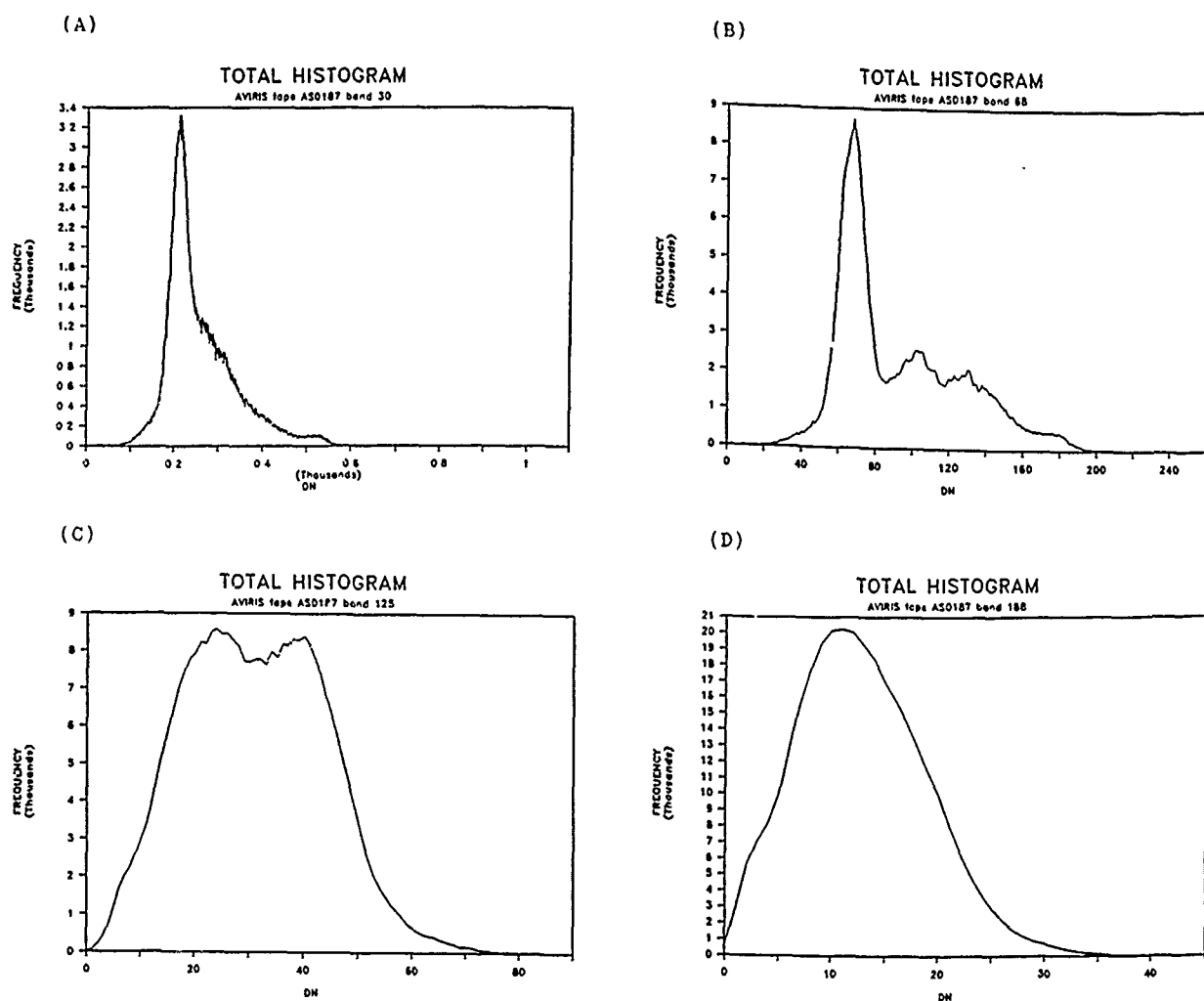


Figure 1. The histograms of the four spectral images of the Cuprite mining district.

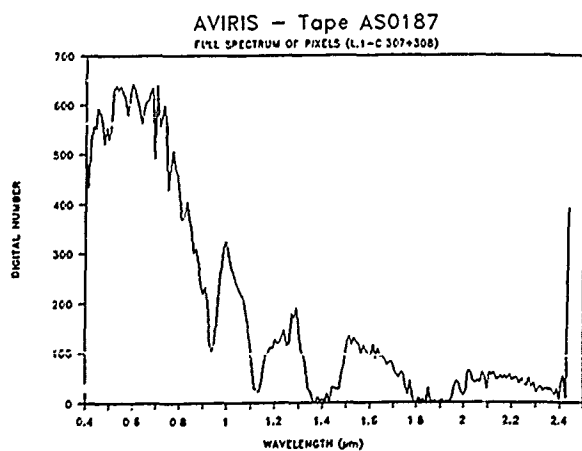


Figure 2. The AVIRIS total spectrum of two adjacent pixels.

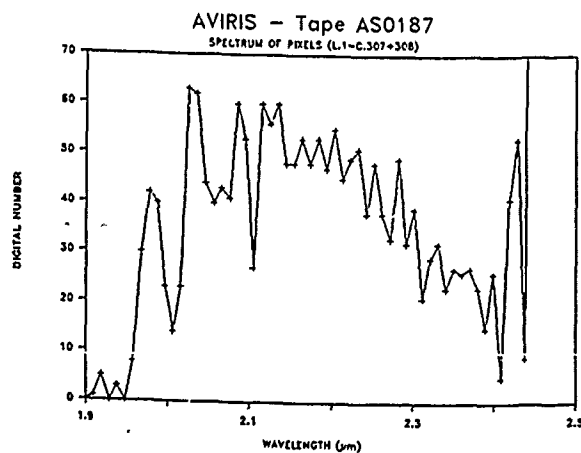


Figure 3. The AVIRIS spectrum between 1.9 and 2.5 μm as in Figure 2.

- Atmospheric model:
midlatitude summer
- Aerosol model:
up to 2 km: rural, 23 km visibility
from 2 to 10 km: tropospheric
from 10 km to 20 km: background strato
- Path:
from 20 km to sea level (angle: 180°)
- Frequency range:
from 4000 to 5250 cm^{-1}

Table 1. Parameters for the LOWTRAN 6 transmittance calculation.

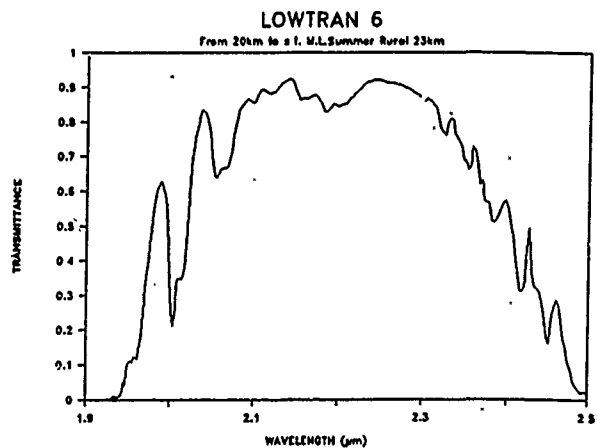


Figure 4. Computed atmospheric transmittance from 1.9 to 2.5 μm .

A STUDY OF THE DOLLY VARDEN MOUNTAINS, NEVADA THROUGH THE USE OF BROAD BAND REMOTE SENSING AND IMAGING SPECTROMETRY

J. A. Zamudio, W. W. Atkinson Jr.*

Center for the Study of Earth from Space (CSES) CIRES, University of Colorado,
Boulder, CO 80309

*Department of Geological Sciences, University of Colorado, Boulder CO 80309

ABSTRACT

Our goal of understanding the Mesozoic tectonic setting of the Dolly Varden Mountains in Nevada was aided by combining remote sensing data with detailed field mapping. Significant faults in the range were delineated by convolving a directional filter with Landsat Thematic Mapper (TM) data. Existing geologic maps were improved upon by analyzing information derived from TM data as well as from the Thermal Infrared Multispectral Scanner (TIMS) and the Airborne Imaging Spectrometer (AIS).

A time constraint is provided by correlating Mesozoic structures with an early Cretaceous intrusion. The association of parts of the stock and its attendant metamorphism to north trending normal faults indicates that in the early to middle Mesozoic the area was undergoing extension along an axis oriented east-west. Early fractures within the stock are filled in some places with aplite, microgranite and other late stage dikes which trend N35°W and suggest that by Cretaceous time the extension axis had rotated to a roughly NE-SW orientation.

Keywords: Imaging Spectrometry, Thematic Mapper, Thermal Infrared Multispectral Scanner, Airborne Imaging Spectrometer

INTRODUCTION

In eastern Nevada and western Utah is an extensive terrane that has experienced a complex tectonic history of Mesozoic deformation and superposed Tertiary extension. In Utah, this terrane is referred to as the Sevier foreland thrust belt where thrusting of sedimentary rocks in an easterly direction occurred during the Cretaceous to early Tertiary Sevier Orogeny. West of the Wasatch Mountains, superposed Cenozoic extension has complicated the tectonic picture. This region, whose western boundary is defined by the thrusts of the Antler orogenic belt (figure 1), is commonly referred to as the hinterland of the Sevier orogenic belt. The Mesozoic history of the region has been controversial for the past twenty or more years (e.g., Misch, 1960; Armstrong, 1972; Miller et al, 1983). Many of the low-angle normal faults of the region were once considered to be thrusts. It is likely that Tertiary extension has reactivated Mesozoic faults (Snow, 1963) further complicating the recognition of

Mesozoic versus Cenozoic structures. A key to a better understanding of the tectonic history of the region is the documentation of geologic field relations coupled with information on the timing of some of the tectonic events.

The Dolly Varden Mountains in eastern Nevada were chosen as the study area for the beginning phase of an investigation of the tectonic history of this part of Nevada because of the amount of previous work already done in the range and the occurrence of an early Cretaceous intrusion to provide some time constraints on the Mesozoic deformation.

Remote sensing data of two general types, broad band and imaging spectrometer, were collected over the area. "Broad band sensors" refers to Landsat TM and TIMS, as their bandwidths are on the order of tens to one thousand nanometers. An imaging spectrometer such as the AIS can collect data in narrow, contiguous bands with widths on the order of ten nanometers. Spectra that are produced from a sensor such as the AIS provide high-spectral-resolution information where the locations and shapes of absorption features can be accurately determined (Goetz et al, 1985). Thus, these data can be used to differentiate between the many mineral species that have various absorption features due to charge transfer, crystal field effects, molecular vibrations and other mechanisms (Goetz et al, 1985). Mineral maps can be produced that show the distribution of the minerals which dominate the picture elements (pixels) in a scene, significantly reducing the amount of field mapping and laboratory analysis required for a geologic study.

GEOLOGY

The Dolly Varden Mountains are located about 60 kilometers southwest of the town of Wendover. The exposed sedimentary section contains 2200 meters of Permian and Triassic miogeoclinal rocks including limestone, dolomite, sandstone, siltstone, chert and shale (Atkinson et al., 1982).

The Mesozoic Melrose Stock which intruded the sedimentary rocks has yielded a K-Ar age of 125 Ma \pm 20, -5 m.y. (Snow, 1963). An adjacent diorite dike was dated at 154 Ma (K-Ar age) by W. J. Moore of the U. S. G. S. (Atkinson et al., 1982). The stock contains at least two phases, a monzonite and a quartz monzonite (Moore, 1976). Metamorphism along the intrusive contact is variable in extent and mineralogy. In the lower part of the Paleozoic section, limestone and sandstone show only

recrystallization in an aureole that locally is less than a meter wide. The degree and extent of metamorphism increases upsection to widths of one kilometer, being most pervasive near some large faults. Andraditic-grossularitic garnet, diopside, wollastonite and minor forsterite are the highest grade minerals found. Subsequent retrograde alteration has produced tremolite, talc, serpentine, montmorillonite, chlorite, sepiolite and saponite. Numerous quartz latite and rhyolite porphyry dikes, possibly related to the stock are found within and near the margin of the intrusion (Atkinson et al., 1982).

Tertiary volcanic rocks of probable Oligocene age cover most of the eastern part of the range (figure 2). They include andesite lavas and quartz latite and rhyolite ash-flow tuffs. These rocks, like the older sedimentary and igneous rocks, have been tilted to the east about 20 degrees as a result of Cenozoic extension.

MESOZOIC STRUCTURES

Folds

Folding in the range is primarily restricted to three anticlines within the Paleozoic section (Snow, 1963; Atkinson et al, 1982). The Victoria Anticline plunges gently ENE and lies just to the south of the Melrose Stock. The stock intruded passively and to some degree the anticline was a favorable structure for emplacement. The Blackhawk Anticline trends north and is slightly overturned with an axial plane that dips to the east. The Keystone Anticline trends northerly and is located on the southeast side of the range. These folds appear to be the oldest structures in the range as they are cut by the second oldest structures, north-trending normal faults (Snow, 1963).

Faults

North-trending normal faults apparently were zones of weakness present during intrusion of the stock, as in some places, parts of the intrusion are sublinearly aligned along them. In the south-central part of the range, calc-silicate metamorphic minerals were formed preferentially along brecciated zones associated with these faults.

Joints and Dikes

The oldest post-intrusive structures are N35°W trending joints in the stock, of which a number have been hydrothermally altered and filled with aplite, pegmatite and microgranite dikes that dip steeply to the southwest (Snow, 1963). This would suggest that immediately after intrusion of the stock, the axis of least principal stress in the region was oriented NE-SW. Fissures on the southeast side of the range appear to be somewhat younger and trend north and northeast. They commonly are filled by veins or altered porphyry dikes.

LANDSAT TM, TIMS AND AIS DATA ANALYSIS

Landsat TM

Directional, three-by-three box filters were convolved in the spatial dimension with band 5 of TM data to enhance linear features in the scene. Three filters were made to enhance features trending north, N17°E and N60°W which correlated with previously mapped faults. This technique delineated mapped major faults and also provided evidence that some faults are more extensive

than previous maps had shown.

Color images were made with various ratio and principal component scenes as well as decorrelation stretches (Moik, 1980; Gillespie et al, 1986). These images delineate the igneous, sedimentary and various volcanic units found in the area. A ratio image with bands 5/7 as red, 3/1 as green and 3/5 as blue shows carbonate rocks as magenta, the monzonite stock as bluish green, rhyolite as lime green, quartz latite as light orange and andesite as purple.

A principal component scene was made putting PC 2 in the red plane, PC 3 in green and PC 4 in blue. Good differentiation of rock units, especially volcanic rocks is evident in this scene.

Finally, a decorrelation stretch was applied using all bands except the thermal one. Three resulting bands that show the most contrast in exposed rocks were then used to make the color image. This image is not as good for differentiation in the carbonate and sandstone sedimentary rocks, but is excellent for the igneous rocks.

All three resulting TM images were used to improve the geologic map of the area and to correlate volcanic units throughout the range. Obviously, field geologists were unable to follow every contact in the field so that delineation of rock units is improved. Differences in the volcanic rocks are especially enhanced in the images; in some cases, these differences are not apparent when comparing adjacent outcrops in the field.

TIMS

TIMS data were enhanced using a decorrelation stretch on bands 1, 3 and 5, then displaying them as blue, green and red respectively. Silica-rich rocks are generally red to orange using this method, and the monzonite, the most silicate-rich rock in the area, is reddish. Good differentiation of felsic through intermediate volcanic rocks is also apparent in the image.

AIS

Airborne Imaging Spectrometer data were collected during the same flight as the TIMS data. The data is concentrated around the metamorphic aureole for the purpose of mapping metamorphic minerals and studying their relationship to faults.

The radiance data acquired by the sensor should be converted to reflectance in order to be able to extract information on the minerals present. As there was no onboard calibration during the flight, the data could not be converted to absolute reflectance. However, the data were converted to internal average relative reflectance using the method described by Kruse (1988). The data were first normalized using an equal energy spectral normalization which scales the sum of the DN's (digital numbers) in each spectrum to a constant value. This step removes albedo differences due to topographic effects. An average spectrum was then calculated for each flight line and divided into each spectrum in that flight line to remove atmospheric and solar effects. One must be aware that any absorption features that are common throughout a majority of pixels will be evident in the average spectrum for the whole flight line. This will have the effect of producing a peak in reflectance for pixels that do not have minerals with that particular feature. These relative (to the average spectrum) reflectance spectra were then

analyzed using the Spectral Analysis Manager (SPAM) program developed by the California Institute of Technology Jet Propulsion Laboratory. In order to handle the large amount of data obtained by an imaging spectrometer, SPAM uses a binary spectral encoding algorithm which can efficiently separate, identify and classify spectra through use of this signature matching method (Goetz et al, 1985). The spectra were compared to library mineral spectra in SPAM (figure 3) and some mineral maps produced showing distribution of such phases as calcite, dolomite, tremolite, kaolinite and montmorillonite.

CONCLUSIONS

Our knowledge of the geologic structure, and distribution of rock types and metamorphic minerals is improved through our use of various remote sensing data types, combined with our field work. Significant faults are easily discerned in directionally filtered TM scenes. It is evident from detailed field mapping and remote sensing data that in pre-Cretaceous time the Dolly Varden area was in an extensional regime as evidenced by the association of parts of the stock and its attendant metamorphism with normal faults. The northerly trends of these faults suggest that the axis of least principal stress was oriented east-west sometime during the early to middle Mesozoic.

The presence of N35°W trending joints in the stock provides evidence that in Cretaceous time the area was still undergoing extension, but the least principal stress direction had rotated to roughly a NE-SW orientation.

ACKNOWLEDGEMENTS

The authors would like to thank JPL for providing SPAM to use in spectral analysis. The study is funded under NASA grant NAGW-1293.

REFERENCES

1. Armstrong R.L., "Low-angle (denudation) faults, hinterland of the Sevier orogenic belt, eastern Nevada and western Utah", Geological Society of America Bulletin, v. 83, pp1729-1754, 1972.
2. Atkinson, W. W., Jr., Kaczmarowski, J. H., and Erickson, A. J., Jr., "Geology of a skarn-breccia ore body at the Victoria mine, Elko County, Nevada", Economic Geology, v. 77, pp899-918, 1982.
3. Gillespie, A. R., Kahle, A. B., and Walker, R. E., "Color enhancement of highly correlated images I. Decorrelation and HSI contrast stretches", Remote Sensing of Environment, v. 20, pp209-235, 1986.
4. Goetz, A. F. H., Vane, G., Solomon, J. E., and Rock, B. N., "Imaging spectrometry for Earth remote sensing", Science, v. 228, pp1147-1153, 1985.
5. Kruse, F. A., "Use of Airborne Imaging Spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada and California", Remote Sensing of Environment, v. 24, no. 1, pp31-51, 1988.
6. Miller, E. L., Gans, P. B., and Garing, J., "The Snake Range decollement: an exhumed mid-Tertiary ductile-brittle transition", Tectonics, v. 2, pp239-263, 1983.
7. Misch, P., "Regional structural reconnaissance in central northeast Nevada and some adjacent areas: Observations and interpretations", Geology of east-Petroleum Geologists 11th Annual Field Conference Guidebook, pp17-42, 1960.
8. Moik, J. G., "Digital processing of remotely sensed data" NASA SP-431, 1980.
9. Moore, T. E., personal communication, 1976.
10. Snow, G. C., "Mineralogy and geology of the Dolly Varden Mountains, Elko County, Nevada", Unpub. Ph.D. thesis, Univ. of Utah, 1963.

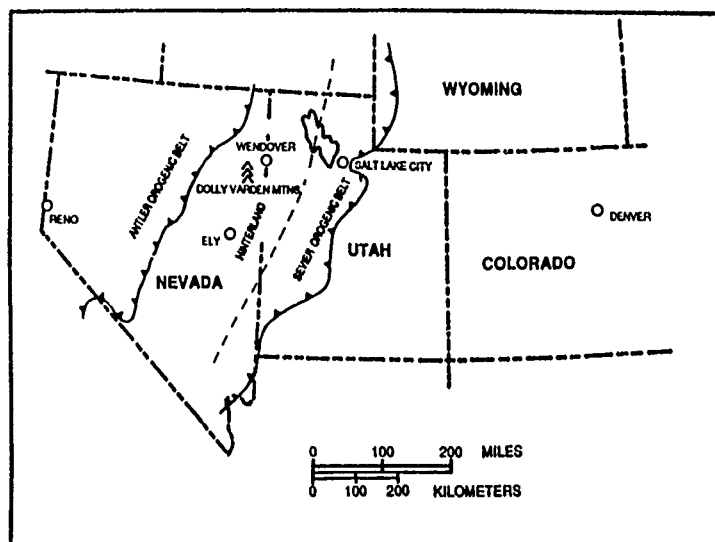


Figure 1. Location of the Dolly Varden Mountains.

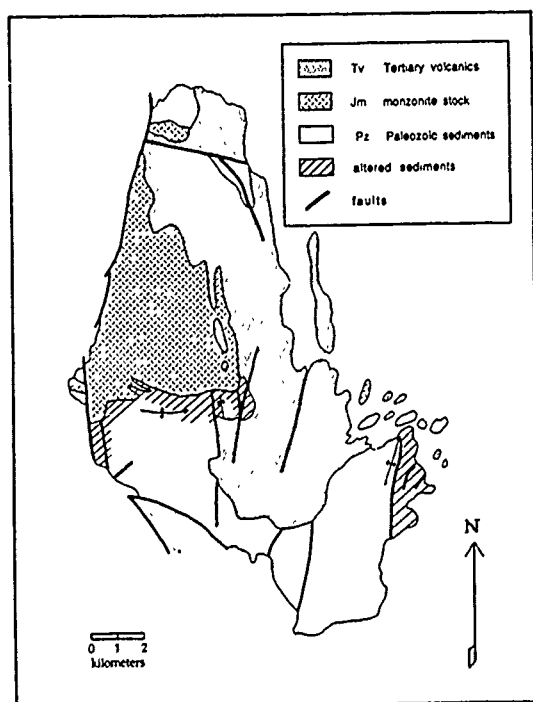


Figure 2. Generalized geologic map of the Dolly Varden Range showing the association of metamorphism and alteration with north trending faults.

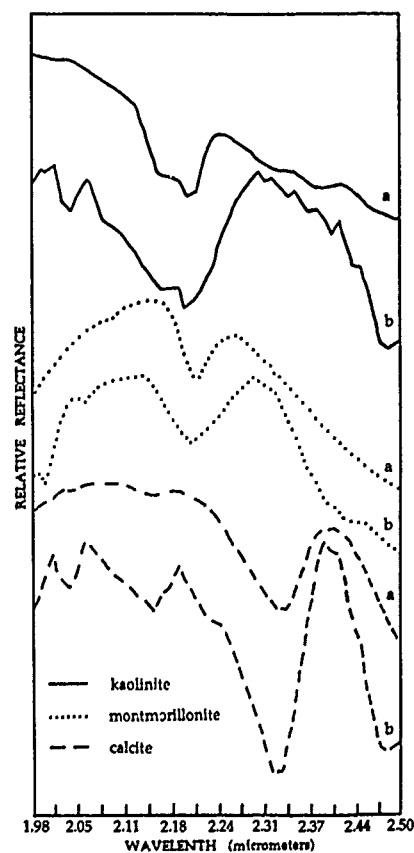


Figure 3. Comparison of: a) laboratory spectra taken by the Portable Instant Display and Analysis Spectrometer (PIDAS). b) spectra taken by the Airborne Imaging Spectrometer (AIS). Spectra are vertically offset for clarity.

MAPPING HYDROTHERMALLY ALTERED ROCKS WITH THE AIRBORNE IMAGING SPECTROMETER (AIS) AND THE AIRBORNE VISIBLE/INFRARED IMAGING SPECTROMETER (AVIRIS)

Fred A. Kruse and Dan L. Taranik
Center for the Study of Earth from Space (CSES)
CIRES, University of Colorado, Boulder, CO 80309
Phone (303) 492-6880, FAX (303)-492-5070

ABSTRACT

The Airborne Imaging Spectrometer (AIS), and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) were used to map the alteration mineralogy of a hydrothermal system in the Northern Grapevine Mountains, Nevada and California using detailed spectral characteristics. Areas of quartz-sericite-pyrite (QSP) alteration were identified based on the presence of sericite (fine-grained muscovite) spectral features near 2.2 μm . Areas of argillic alteration were defined based on the presence of montmorillonite. Calcite and dolomite were identified using the AIS based on sharp absorption features near 2.3 μm . Iron oxide distribution was mapped using the AVIRIS visible wavelengths. The mineralization observed is similar to that seen in some disseminated porphyry copper type deposits. The imaging spectrometer data, however, show that the distribution of alteration is usually fracture controlled and field checking indicates that widespread supergene alteration is not present.

Keywords: Imaging spectrometry, mineral mapping, hydrothermal alteration

INTRODUCTION

Detailed maps of the distribution of alteration minerals at the surface are often the key to understanding the processes by which mineral deposits form and to determining the location of ore bodies. These maps are usually produced by detailed field mapping combined with expensive laboratory analysis techniques. High spectral resolution remote sensing (imaging spectrometry) is a new tool that can be used to quickly produce detailed maps for previously unmapped areas, and to guide field geologic mapping efforts. It can substantially reduce required field mapping and laboratory analyses by making positive identification of mineralogy possible.

Imaging spectrometry is "the simultaneous acquisition of images in many narrow, contiguous spectral bands" (Goetz et al, 1985). Analysis of imaging spectrometer data allows extraction of a detailed spectrum for each picture element (pixel) of the image. The AIS was an experimental imaging spectrometer flown from 1983-87 to test two-dimensional, near-infrared area array detectors. This instrument imaged 32 (AIS-1) or 64 (AIS-2) cross-track pixels simultaneously, collecting data in 128 contiguous narrow (9.3 nm) channels from approximately 1.2 to 2.4 μm (Vane et al, 1983). The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is a 224-channel instrument measuring surface radiance over the spectral range 0.41 to 2.45 μm in approximately 10 nm-wide bands (Porter and Enmark, 1987).

High spectral resolution reflectance spectra such as those

collected by imaging spectrometers allow direct identification of minerals based upon their reflectance characteristics. Molecular vibrations in minerals result in characteristic overtone and combination tone absorption bands in the infrared near 1.4, 1.9, 2.2, and 2.3 μm (Hunt, 1977). The positions and shapes of these absorption features vary in a predictable fashion for different minerals (Hunt and Salisbury, 1970; Hunt et al, 1971; Lee and Raines, 1984). Additional broad absorption bands are present in minerals such as iron oxides at wavelengths less than about 0.95 μm because of intervalence charge transfer between O^{2-} and Fe^{3+} and crystal field transitions dependent on the type and degree of crystallinity (Hunt, 1977).

GEOLOGY

The study area in the northern Grapevine Mountains, Nevada, has been studied in detail using conventional geologic mapping, geochemistry, field and laboratory reflectance spectroscopy, and imaging spectrometers (Wrucke et al, 1984; Kruse, 1988). Precambrian bedrock consists of limestones, dolomites, sandstones and their metamorphic equivalents. Mesozoic plutonic rocks include quartz syenite, a quartz monzonite porphyry stock, and quartz monzonite dikes. Tertiary volcanic rocks are abundant around the periphery of the study area (Wrucke et al, 1984). Quaternary deposits include Holocene and Pleistocene fanglomerates, pediment gravels, and alluvium.

The Mesozoic rocks are cut by narrow north-trending mineralized shear zones containing sericite (fine grained muscovite) and iron oxide minerals. Slightly broader zones of disseminated quartz, pyrite, sericite, chalcopyrite, and fluorite mineralization occur in the quartz monzonite porphyry. This type of alteration is spatially associated with fine-grained quartz monzonite dikes. There are several small areas of quartz stockwork exposed at the surface in the center of the area. Skarn, composed mainly of brown andradite garnet intergrown with calcite, epidote, and tremolite, occurs around the perimeter of the quartz monzonite stock in Precambrian rocks.

IMAGING SPECTROMETER DATA ANALYSIS

Preprocessing

Preprocessing of imaging spectrometer data is usually required to remove both dropped lines and bad bands from the data. Individual bad lines are replaced with the average of the two adjacent lines. In the AIS data, a pronounced vertical striping pattern was also observed in the images, probably caused by varying dark current offsets in detectors in the pixel direction. This striping was removed using a histogram-matching algorithm (Dykstra and Segal, 1985). Bad bands in both the AIS and AVIRIS data were replaced with the average

of adjacent bands. The dark current file was subtracted from the AVIRIS data. A running 7 line by 1 band box was used to filter the data before subtraction because of the high noise level in the AVIRIS dark current data.

Wavelength Calibration

Laboratory measurements by Jet Propulsion Laboratory (JPL) provided the initial wavelength calibration for both the AIS and AVIRIS data. An additional check on the wavelength calibration was made by comparing the positions of known atmospheric absorption features to their locations in the imaging spectrometer data. Atmospheric carbon dioxide (CO₂) absorption band, located at 2.005 and 2.055 μm are useful for wavelength-calibration of the data in the infrared (Vane, 1987). Comparison of the positions of these band in both the AIS and AVIRIS data showed that the JPL calibrations were accurate to within one channel.

Calibration to Reflectance

Calibration to reflectance is mandatory for detailed analysis of imaging spectrometer data. Ideally the data should be calibrated to absolute reflectance. This requires onboard calibration for each flight which was not available for the imaging spectrometer data. Both the AIS and AVIRIS data used in this study were initially converted to internal average relative (IAR) reflectance (Kruse, 1988). This conversion was selected because it does not require *a priori* knowledge of the site. The IAR reflectance procedure requires that albedo differences be removed from the images so that the spectral information can be extracted. This was accomplished using an "equal area normalization" as described in Vane and Goetz (1985). The normalization scales the sum of the DN's in each spectrum (each pixel with N bands) to a constant value. IAR reflectance is then calculated by determining an average spectrum for all flightlines acquired on an individual mission and dividing each spectrum in each flightline by the average spectrum. The resulting spectra represent reflectance relative to the average spectrum and resemble laboratory spectra acquired of the same materials (Kruse, 1988). When looking at an IAR reflectance spectrum it should be remembered that the average spectrum used to calculate the IAR reflectance spectrum may itself have spectral character related to mineral absorption features. This can adversely affect the appearance of the IAR reflectance spectra and limit their usefulness in comparisons with laboratory spectra (Clark et al, 1987). The average spectrum for the flightlines analyzed in this study did not contain any obvious mineral absorption features.

An alternative to onboard calibration or IAR reflectance conversion is to use a standard area on the ground to calibrate the data (Roberts et al, 1985) however, this approach requires *a priori* knowledge of each site. The calibration to reflectance requires choosing two ground target regions with albedos that span a wide range and acquiring field spectra to characterize them. Field spectra were measured and samples were collected for the study area during 1984-1988. For this study the two calibration targets used were volcanic tuff (bright target) and quartz syenite outcrops and gravel (dark target). The second step in the process involves picking the multiple pixels in the airborne data set that are associated with each ground target. Then a linear regression is calculated for each band to determine the gains and offsets required to convert the DN to reflectance (Kruse et al, 1989). The final step in the calibration is to multiply the instrument DN values by the proper gain factor and to add the corresponding offset value. The result is the removal of atmospheric effects (both attenuation and scattering), viewing geometry effects, and residual instrument artifacts. While no such correction can be perfect, it does allow conversion of the remotely sensed spectra into a form that can be readily compared with laboratory or field acquired spectra.

Analysis techniques

The last 32 channels (2.1-2.4 μm) of the AIS data in the IAR reflectance format were analyzed using automated absorption band mapping techniques (Kruse, 1988). The strongest absorber feature in the 2.1-2.4 μm portion of each AIS spectrum was mapped into an intensity, hue, saturation (IHS) color transform (Raines, 1977) and then transformed into the red, green, blue (RGB) color space. The band position was mapped into hue, the band depth into intensity, and the band width into saturation. Transformation of the IHS-encoded spectral information into the RGB color space produced an "IHS absorption band image" in which all the absorption band information for the strongest absorption feature between 2.1 and 2.4 μm in each pixel was present in the color variation (Kruse, 1988). A mineralogical map was then produced through visual interpretation of the color image and examination of individual AIS spectra.

The AVIRIS data were analyzed using a new software package called "Integrated Software for Imaging Spectrometers" (ISIS) (Torson, 1989) developed by the United States Geological Survey in Flagstaff, Arizona and enhanced by CSES/University of Colorado. ISIS is a software package that allows interactive analysis of imaging spectrometer data. Individual pixels and area averages can be selected for spectral analysis. Part of the image display shows a spatial image and the rest of the display shows a stacked, gray-scaled or color-coded spectral image (Marsh and McKeon, 1983; Kruse, 1988). The stacked spectra correspond to a slice through the spatial image which can be interactively selected in real time. Concurrent tasks allow spectral matching of image spectra to library spectra and overlay of the matched areas onto a single band image.

AIS RESULTS

Two AIS-1 flightlines acquired in 1984, one AIS-1 flightline acquired in 1985, and four AIS-2 flightlines acquired in 1986 were prepared as IHS-coded absorption band images for spectral analysis. Individual spectra and groups of spectra were extracted from the AIS data and comparisons were made to laboratory standards and published spectra to identify alteration minerals (Hunt, 1979; Hunt and Ashley, 1979; Lee and Raines, 1984). Several minerals were identified using the AIS data. Some of the alteration in the area consists of quartz-sericite-pyrite (QSP) alteration as identified from absorption features at 2.207, 2.245, and 2.346 μm for sericite (fine-grained muscovite) (Figure 1). Areas of argillic alteration were defined by the presence of a weak to moderate absorption feature near 2.21 μm caused by montmorillonite and the absence of the 2.346 μm sericite band (Figure 1). Areas of mixed montmorillonite and sericite could not be distinguished from those having only sericite. The AIS data also allowed positive identification of calcite and dolomite based on the presence of absorption features near 2.34 and 2.32 μm , respectively (Figure 2) and zeolite-group minerals based on an absorption band near 2.4 μm (Kruse, 1988).

AVIRIS RESULTS

The AVIRIS data for this site were reduced to IAR reflectance as described in the preprocessing section. The ISIS program was used to extract spectra for areas of known mineralogy. Figure 3 shows an AVIRIS spectrum for sericite (fine grained muscovite) compared to a laboratory spectrum of muscovite. Figure 4 shows an AVIRIS carbonate spectrum compared to a laboratory spectrum of dolomite. Both AVIRIS spectra are very noisy, although they are averages of several pixels. Severe signal-to noise problems degraded the quality of all spectra extracted from the AVIRIS data and differentiation between calcite and limestone, and muscovite and montmorillonite, previously demonstrated with AIS data, was not possible using the AVIRIS spectra. These findings indicate

that the AVIRIS data is not as useful as the AIS data for mapping subtle mineralogical variation, primarily because of low signal-to-noise. Although the AVIRIS data were useful for mapping strong mineral absorption features and producing mineral maps at the Nevada site, it is clear that significant improvements to the instrument performance are required before AVIRIS will be an operational instrument.

An additional attempt to map minerals was made after the calibration to reflectance using ground targets. Spectra, profiles, and stacked, color-coded spectra were extracted from the AVIRIS data and these derivative data were compared to AIS results, field and laboratory spectra, and geologic maps. Average spectra were extracted for two types of carbonate spectra and one typical "clay" spectrum. These spectra appear similar to the IAR reflectance spectra. Even though these minerals have very strong absorption features, they are only marginally identifiable as mineral groups using the AVIRIS data because of high noise in the data. Binary encoding (Mazer et al, 1988) was used to map the distribution of these the carbonates and clay. Additionally, a representative iron oxide spectrum was extracted from the data (Figure 5) and used to map surface iron oxide distribution.

CONCLUSIONS

The mineral assemblages and distributions mapped from the AIS, and AVIRIS data were used to assist in development of a model for the alteration and mineralization for the study area. The two alteration types (QSP and argillic alteration) defined with the imaging spectrometer data appear to be related to separate stages of mineralization. Field checking shows that quartz-sericite-pyrite (QSP) alteration is the predominant alteration type in the area and is spatially associated with quartz monzonite dikes and possibly a larger quartz monzonite body at depth. New areas of quartz-sericite-pyrite (QSP) alteration were identified using the remote sensing data. The argillic alteration zone occurs in quartz syenite, and is spatially associated with a granitic intrusion. The mineralization observed at the northern Grapevine Mountains site is similar to that seen in some disseminated porphyry copper type deposits, however, it is localized, mostly fracture controlled, and widespread supergene alteration is not present.

REFERENCES CITED

- Clark, R. N., King, T. V. V., and Gorelick, N. S., 1987, Automatic continuum analysis of reflectance spectra: in *Proceedings, Third AIS workshop*, 2-4 June, 1987, JPL Publication 87-30, Jet Propulsion Laboratory, Pasadena, California, p. 138-142.
- Dykstra, J. D., and Segal, D. B., 1985, Analysis of AIS data of the Recluse Oil Field, Recluse, Wyoming: in *Proceedings, AIS workshop*, 8-10 April, 1985, JPL Publication 85-41, Jet Propulsion Laboratory, Pasadena, California, p. 86-91.
- Goetz, A. F. H., Vane, G., Solomon, J. E., and Rock, B. N., 1985, *Imaging spectrometry for earth remote sensing: Science*, v. 228, p. 1147-1153.
- Hunt, G. R., 1977, Spectral signatures of particulate minerals in the visible and near-infrared: *Geophysics*, v. 42, no. 3, p. 501-513.
- Hunt, G. R., 1979, Near-infrared (1.3-2.4 μ m) spectra of alteration minerals - potential for use in remote sensing: *Geophysics*, v. 44, p. 1974-1986.
- Hunt, G. R., and Ashley, R. P., 1979, Spectra of altered rocks in the visible and near infrared: *Econ. Geol.*, v. 74, p. 1613-1629.
- Hunt, G. R., and Salisbury, J. W., 1970, Visible and near-infrared spectra of minerals and rocks: II. Carbonates: *Mod. Geol.*, v. 2, p. 23-30.
- Hunt, G. R., Salisbury, J. W., and Lenhof, C. J., 1971, Visible and near-infrared spectra of minerals and rocks: III. Oxides and hydroxides: *Mod. Geol.*, v. 2, p. 195-205.
- Kruse, F. A., 1988, Use of Airborne Imaging Spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada and California: *Remote Sensing of Environment*, v. 24, no. 1, p. 31-51.
- Kruse, F. A., Kierein-Young, K. S., and Boardman, J. W., 1989, Mineral mapping at Cuprite, Nevada with a 63 channel imaging spectrometer: *Photogrammetric Engineering and Remote Sensing*, v. LV, no. 4, September, 1989 (submitted).
- Lee, Keenan, and Raines, G. L., 1984, Reflectance spectra of some alteration minerals--a chart compiled from published data 0.4 μ m-2.5 μ m: *U.S. Geological Survey Open-File Report 84-96*, 6 p., 1 chart.
- Marsh, S. E., and McKeon, J. B., 1983, Integrated analysis of high-resolution field and airborne spectroradiometer data for alteration mapping: *Econ. Geol.*, v. 78, no. 4, p. 618-632.
- Mazer, A. S., Martin, M., Lee, M., and Solomon, J. E., 1988, Image processing software for imaging spectrometry data analysis: *Remote Sensing of Environment*, v. 24, no. 1, p. 201-210.
- Porter, W. M., and Enmark, H. T., 1987, A system overview of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): in *Proceedings, 31st Annual International Technical Symposium*, 16-21 August, 1987, Society of Photo-Optical Instrumentation Engineers, V. 834, p. 22-31.
- Raines, G. R., 1977, Digital color analysis of color ratio composite Landsat scenes: in *Proceedings, Eleventh International Symposium on Remote Sensing of Environment*, University of Michigan, Ann Arbor, p. 1463-1472.
- Roberts, D. A., Yamaguchi, Y., and Lyon, R. J. P., 1985, Calibration of Airborne Imaging Spectrometer Data to percent reflectance using field spectral measurements: in *Proceedings, Nineteenth International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, October 21-25, 1985.
- Torson, J. M., 1989, Interactive image cube visualization and analysis: in *Proceedings, Chapel Hill Workshop on Volume Visualization*, 18-19 May, 1989, University of North Carolina at Chapel Hill, (in press).
- Vane, Gregg, 1987, First results from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): in *Proceedings, 31st Annual International Technical Symposium*, 16-21 August, 1987, Society of Photo-Optical Instrumentation Engineers, V. 834, p. 166-174.
- Vane, Gregg, and Goetz, A. F. H., 1985, (eds.), *Proceedings of the Airborne Imaging Spectrometer (AIS) Data Analysis Workshop*, 8-10 April 1985, JPL Publication 85-41, Jet Propulsion Laboratory, Pasadena, California, 173 p.
- Vane, Gregg, Goetz, A. F. H., and Wellman, J. B., 1983, Airborne imaging spectrometer: A New Tool for Remote Sensing: *IEEE Transactions on Geoscience and Remote Sensing*, v. GE-22, no. 6, p. 546-549.
- Wrucke, C. T., Werschky, R. S., Raines, G. L., Blakely, R. J., Hoover, D. B., and Miller, M. S., 1984, Mineral resources and mineral resource potential of the Little Sand Spring Wilderness Study Area, Inyo County, California: *U.S. Geological Survey Open File Report 84-557*, 20 p.

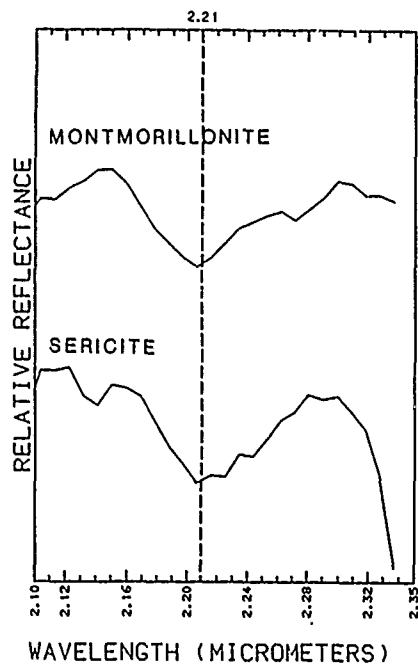


Figure 1. AIS-1 spectrum for montmorillonite compared to an AIS-1 spectrum for sericite. Note that the sericite spectrum drops off towards a 2.35 μm absorption feature while montmorillonite spectrum does not. Spectra are offset for clarity.

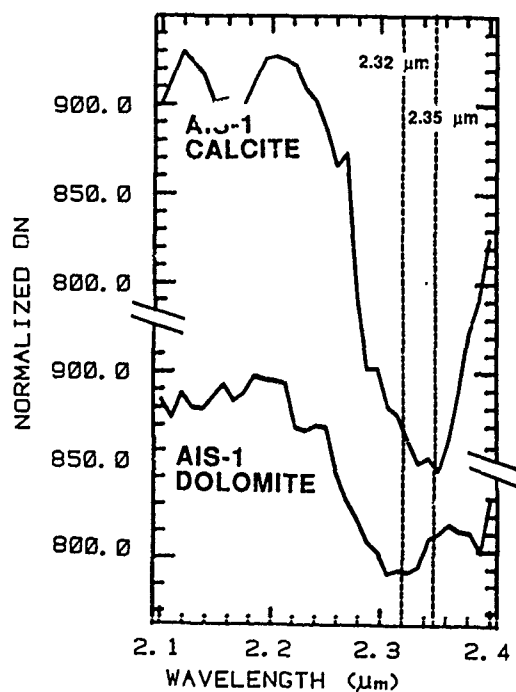


Figure 2. AIS-1 spectrum of calcite (2.35 μm) compared to AIS-1 spectrum of dolomite (2.32 μm). Spectra are offset for clarity.

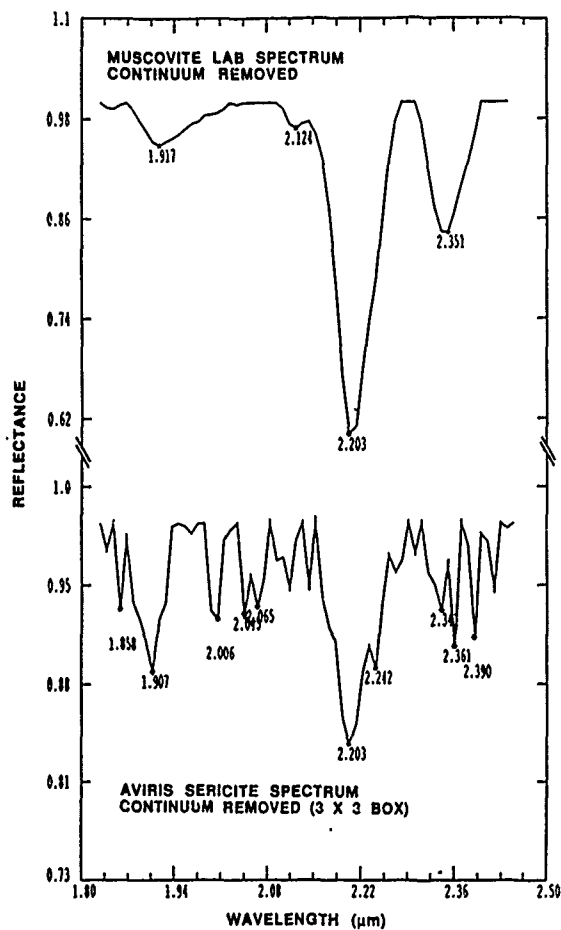


Figure 3. Comparison of continuum removed AVIRIS spectrum of sericite and lab spectrum of muscovite. Spectra are offset for clarity.

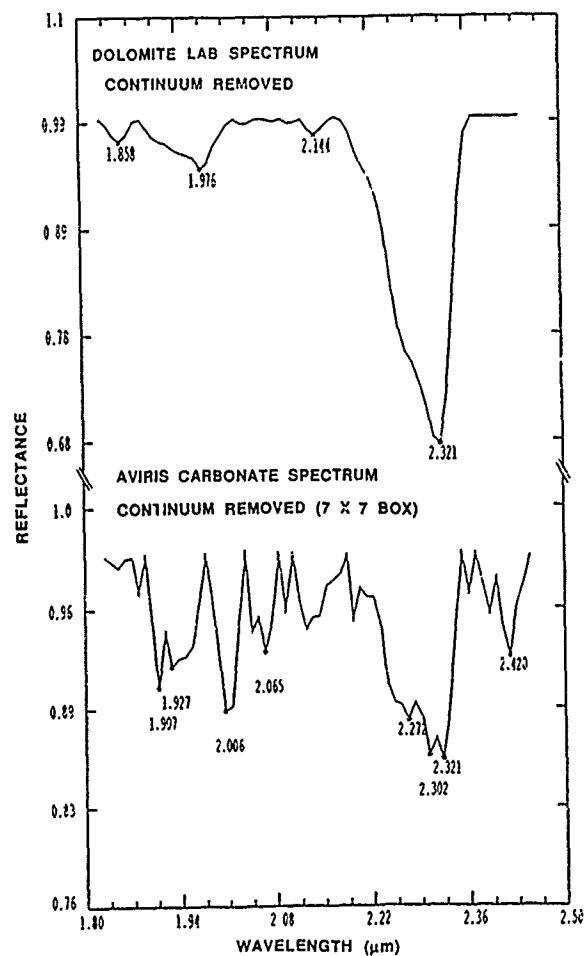


Figure 4. Comparison of continuum removed AVIRIS carbonate spectrum to dolomite. Spectra are offset for clarity.

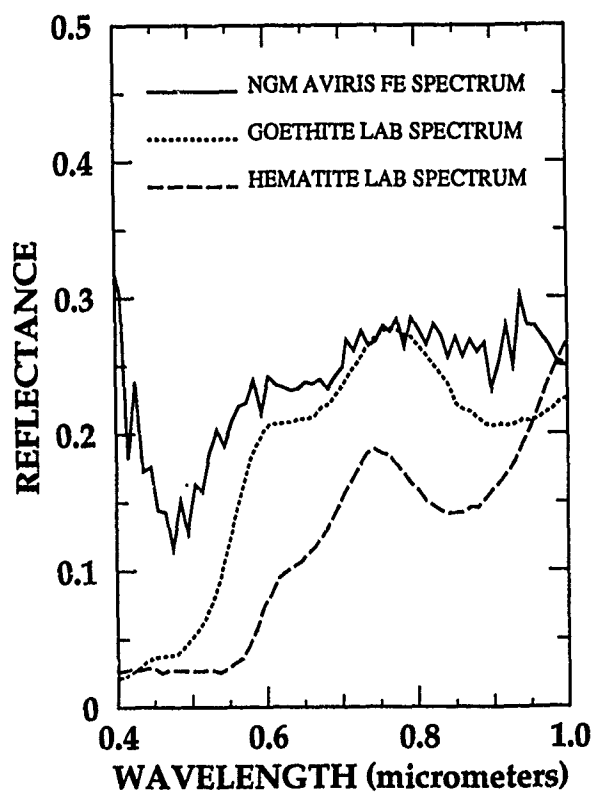


Figure 5. Comparison of AVIRIS iron oxide spectrum to hematite and goethite lab spectra. Poor signal-to-noise characteristics make identification difficult, but the AVIRIS spectrum more closely resembles goethite than hematite.

LITHOLOGY AND STRUCTURE WITHIN THE BASEMENT TERRAIN ADJACENT TO
CLARK MOUNTAINS, CALIFORNIA, MAPPED WITH CALIBRATED DATA FROM THE
AIRBORNE VISIBLE/INFRARED IMAGING SPECTROMETER

Robert O. Green and Gregg Vane

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

BACKGROUND

The Clark Mountains lie just to the north of Interstate 15 in eastern California, between Barstow, California and Las Vegas, Nevada. The rugged, highly dissected area is nearly 5000 feet above sea level, with Clark Mountain rising to 8000 feet. The rocks comprising the Clark Mountains and the Mescal Range just to the south are Paleozoic carbonate and clastic rocks, and Mesozoic clastic and volcanic rocks standing in pronounced relief above the fractured Precambrian gneisses to the east (Evans, 1974). The Permian Kaibab Limestone and the Triassic Moenkopi and Chinle Formations are exposed in the Mescal Range, which is the only place in California where these rocks, which are typical of the Colorado Plateau, are found. To the west, the mountains are bordered by the broad alluvial plains of Shadow Valley. Cima Dome, which is an erosional remnant carved on a batholithic intrusion of quartz monzonite, is found at the south end of the valley. To the east of the Clark and Mescal Mountains is found the Ivanpah Valley, in the center of which is located the Ivanpah Playa. The geography of the area is shown in Figure 1.

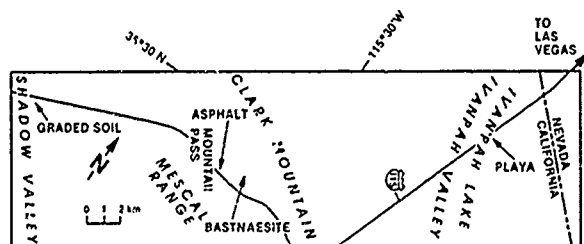


Figure 1. Location of the AVIRIS imagery acquired over the Clark Mountains in 1987. Shown are Ivanpah Valley, the Clark Mountains, the Mescal Range, and Shadow Valley. The locations of several calibration targets used in the study are also indicated.

The Clark Mountains are best known for the Mountain Pass carbonatite complex, which contains the world's largest known concentration of rare earth elements in an area that is about 7 miles long by 3 miles wide (Olson et al., 1954). In a 15 mile radius around this area there have been many mines and prospects for gold and copper, as well as lead, zinc, tungsten, tin, silver and antimony. Unrelated to the deposits of rare earths, the metallic mineralization is controlled largely by faults and intrusives of Cretaceous age. The Mountain Pass rare earth district lies in a block of Precambrian rock bounded on three sides by faults and on the fourth by the alluvium of the Ivanpah Valley. The country rock consists of a complex of gneisses and schists. Cutting across the foliation of the metamorphic complex are bodies of alkalic rock, associated with which are hundreds of dikes of carbonate and a large carbonatite body. There are also swarms of pegmatite dikes trending northward, and swarms of andesite dikes of younger age trending eastward. The rare earth mineralization is restricted to the large carbonatite body and carbonate dikes and occurs primarily as the mineral bastnaesite (Olson and Pray, 1954). The diverse mineralogical composition of the Mountain Pass area make it an attractive target for research in high spectral resolution remote sensing.

STUDIES OF THE CLARK MOUNTAINS WITH THE AIRBORNE VISIBLE/INFRARED IMAGING SPECTROMETER

In 1987, images were acquired of the Clark Mountains with the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) during its first flight season. AVIRIS is a whisk-broom sensor that acquires images in 220 contiguous 10 nm wide spectral bands over a 10.5 km swath with a ground instantaneous field of view of 20 m (Vane, 1987). The purpose of the Clark Mountains

experiment was dual: To study the geology of the area, and to assess the performance of the sensor in a complex geological setting. The latter was the primary purpose of the 1987 experiment, and will be addressed in summary fashion in this paper. A subsequent experiment is being conducted in 1989 which will address the geological objectives. One of the chief attractions of the Clark Range from a sensor performance evaluation standpoint is the rare-earth-bearing carbonatite of Mountain Pass. As noted above, the Mountain Pass carbonatite contains a high concentration of the mineral bastnaesite, which has seven generally recognized rare earth metals, although others are present in very small amounts. The seven are cerium, lanthanum, neodymium, praseodymium, samarium, gadolinium, and europium. Calcite is also a major constituent of bastnaesite. Laboratory spectral reflectance curves for 4 of the bastnaesite rare-earth-element oxides are shown in Figure 2 (Rowan et al., 1986). In Figure 3, a laboratory spectral reflectance curve for bastnaesite itself is shown (Adams, 1965). The abundant absorption features result from electronic transitions in the plus-three-charged rare earth elements. These features in conjunction with various narrow atmospheric absorption features were used to assess the spectral resolution of AVIRIS.

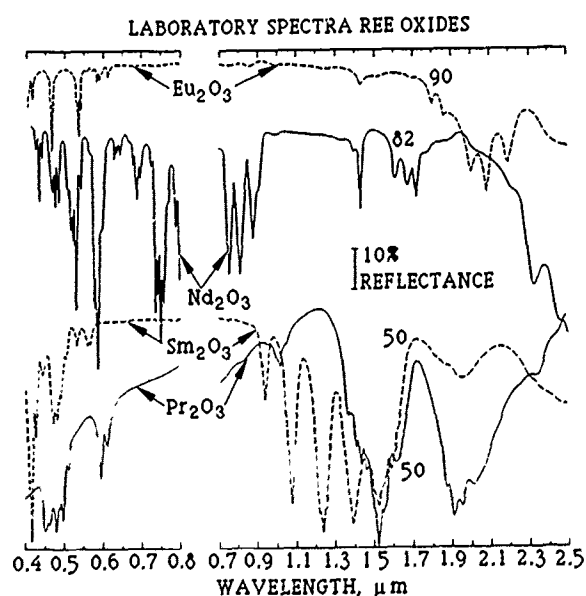


Figure 2. Laboratory reflectance spectra for the four rare-earth oxides of europium, neodymium, samarium and praseodymium, all of which occur in the mineral bastnaesite, one of the major components of the Mountain Pass carbonatite (from Rowan et al., 1986).

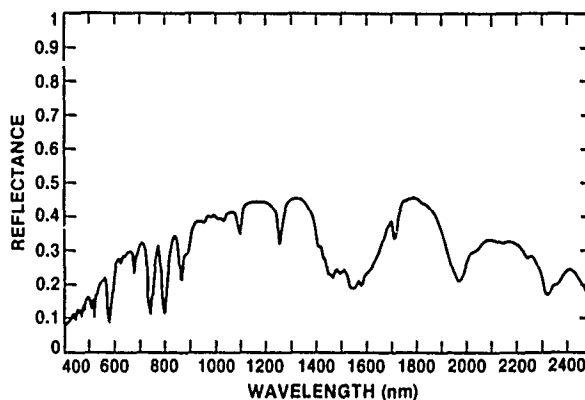


Figure 3. A reflectance spectrum of the mineral bastnaesite acquired with a Beckman UV5240 laboratory spectroradiometer showing the sharp absorption features in the visible and near infrared spectral regions caused by trivalent rare earth elements (Adams, 1965).

An AVIRIS image of Mountain Pass is shown in Figure 4. The area of active mining is immediately north of the highway crossing the image. Figure 5 is an AVIRIS spectrum from a pixel within the open mine pit (Green et al., 1988). The spectrum has been corrected to reflectance using the empirical line algorithm described by Conel et al. (1987), which compensates for multiplicative components of solar illumination, atmospheric attenuation and instrument response, as well as for the additive factors of atmospheric path radiance and instrument dark current. The gaps in the spectrum centered at approximately 940, 1125, 1400 and 1900 nm are due to strong atmospheric water absorption. The offsets in the reflectance curves at about 700 and 1280 nm are due to changes in the response functions of the AVIRIS spectrometers over the duration of the flight line (Vane et al., 1988). Three strong absorption features are clearly seen between 700 and 900 nm. These are due to the presence of neodymium. A smaller feature can also be seen at 600 nm which is due to neodymium or praseodymium, or a combination of both. Other absorption features evident in Figure 5 include a possible neodymium feature at about 1600 nm and a strong CO₃ absorption at 2300 nm. Possible rare earth absorption feature associated with europium may be present between 2000 and 2100 nm, but low signal-to-noise performance in this part of the spectrum preclude drawing strong conclusions about these low amplitude, high resolution features.

Work with this and other data sets from



Figure 4. An AVIRIS image in one 10-nm wide spectral band centered at about 1013 nm. The area imaged is about 10 by 11 km. The Mountain Pass mine is just to the north of Interstate 15.

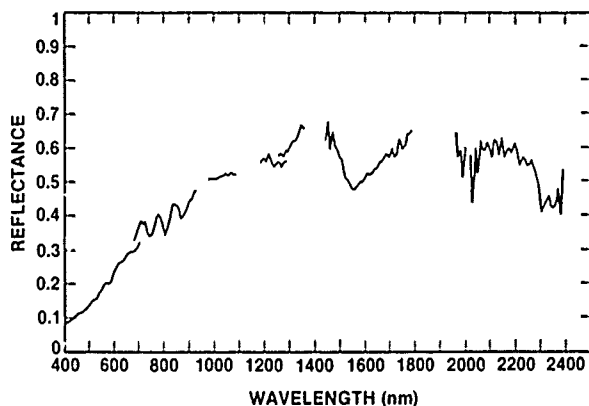


Figure 5. An AVIRIS reflectance spectrum from a pixel centered at the open pit mine at Mountain Pass (Green et al., 1988). The spectrum has been corrected using the empirical line reflectance retrieval algorithm. The three strong neodymium absorption features between 700 and 900 nm are well resolved. Other features in the spectrum are discussed in the text.

1987 confirmed that AVIRIS met many of its performance requirements, but deficiencies were found in the signal-to-noise and in the spectral resolution of some of the AVIRIS spectrometers. The radiometric response function was also found to vary from flight to flight and to a lesser degree, within a given flight. Additional engineering work was done on the sensor in 1988. Results will be presented at the symposium on the performance of the sensor based on recent in-flight experiments over the Clark Mountains. Also to be presented will be structural and lithological mapping based on the analysis of calibrated AVIRIS data from the Clark Mountains.

ACKNOWLEDGEMENTS

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

- Adams, J.W., "The visible region absorption spectra of rare-earth minerals," *The Amer. Min.*, Vol. 50, pp 356-367, 1965.
- Conel, J.E., Green, R.O., Vane, G., Bruegge, C.J., and Alley, R.E., "AIS-2 radiometry and a comparison of methods for the recovery of ground reflectance," in *Proceedings of the Third Airborne Imaging Spectrometer Data Analysis Workshop*, G. Vane, ed., JPL Publication 87-30, pp 18-47, 1987.
- Evans, J.R., "Relationship of mineralization to major structural features in the Mountain Pass area, San Bernardino County, California," *California Geology*, pp 516-526, July 1974.
- Green, R.O., Vane, G., and Conel, J.E., "Determination of . . . light AVIRIS spectral, radiometric, spatial and signal-to-noise characteristics using atmospheric and surface measurements from the vicinity of the rare-earth-bearing carbonatite at Mountain Pass, California," in *Proceedings of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Performance Evaluation Workshop*, G. Vane, ed., JPL Publication 88-38, pp 162-184, 1988.
- Olson, J.C. and Pray, L.C., "The Mountain Pass rare earth deposits," *Calif. Div. Mines and Geol., Bull.* 170, pp 23-29, 1954.
- Olson, J.C., Shawe, D.R., Pray, L.C., and Sharp, W.N., "Rare earth mineral deposits of the Mountain Pass district, San Bernardino County, California," *U.S. Geol. Survey Prof. Paper*, Vol. 261, 75 pp, 1954.
- Rowan, L.C., Kingston, M.J., and Crowley, J.K., "Spectral reflectance of carbonatites and related alkalic igneous rocks: Selected samples from four North American localities," *Econ. Geol.*, Vol. 81, pp 857-871, 1986.
- Vane, G., ed., "Airborne Visible/Infrared Imaging Spectrometer (AVIRIS): A description of the sensor, ground data processing facility, laboratory calibration, and first results," JPL Publication 87-38, 97 pp, 1987.
- Vane, G., Chrien, T.G., Reimer, J.H., Green, R.O., and Conel, J.E., "Comparison of laboratory calibrations of the Airborne Visible/Infrared Imaging Spectrometer at the beginning and end of the first flight season," in *Proceedings of the SPIE Conference on Recent Advances on Sensors, Radiometry and Data Processing for Remote Sensing*, P. Slater, ed., Vol. 924, pp 168-178, 1988.

COMPARISON OF LANDSAT THEMATIC MAPPER AND GEOPHYSICAL AND ENVIRONMENTAL RESEARCH IMAGING SPECTROMETER DATA FOR THE CUPRITE MINING DISTRICT, ESMERALDA AND NYE COUNTIES, NEVADA

Kathryn S. Kierein-Young and Fred A. Kruse

Center for the Study of Earth from Space (CSES)
CIRES, University of Colorado, Boulder, CO 80309
Phone (303) 492-5130, FAX (303) 492-5070

ABSTRACT

LANDSAT Thematic Mapper (TM) images and Geophysical and Environmental Research Imaging Spectrometer (GERIS) data were analyzed for the Cuprite mining district and compared to available geologic and alteration maps of the area. The TM data with 30 meter resolution and 6 broad bands allowed discrimination of general mineral groups. Clay minerals, playa deposits, and unaltered rocks were mapped as discrete spectral units using the TM data, but specific minerals were not determined and definition of the individual alteration zones was not possible.

The GERIS, with 15 meter spatial resolution and 63 spectral bands, permitted construction of complete spectra and identification of specific minerals. Detailed spectra extracted from the images provided the ability to identify the minerals alunite, kaolinite, hematite, and buddingtonite in the Cuprite district by their spectral characteristics. The GERIS data show a roughly concentrically zoned hydrothermal system. The mineralogy mapped with the aircraft system conforms to previous field and multispectral image mapping. However, identification of individual minerals and spatial display of the dominant mineralogy add information that can be used to help determine the morphology and genetic origin of the hydrothermal system.

Keywords: Imaging Spectrometry, Thematic Mapper, Cuprite, Nevada

INTRODUCTION

The Cuprite mining district, located about 15 km south of Goldfield on U.S. Highway 95 in southwest Nevada (Figure 1), is an excellent area to test remote sensing technology because of the good rock exposures and the presence of several distinct mineral assemblages. Cuprite has been used for many studies over the years and an extensive image database and collection of field and laboratory spectra exist for the district (Rowan et al, 1974; Kahle and Goetz, 1983; Goetz and Srivastava, 1985). The geology of the district is relatively well known and has been described in detail by Abrams et al (1977), Ashley and Abrams (1980), and Shipman and Adams (1987). Bedrock consists of Tertiary volcanic and volcanoclastic rocks, principally rhyolitic ash-flow tuffs with some air-fall tuff (Abrams et al, 1977). The volcanic rocks have been extensively modified in the Cuprite district by hydrothermal

alteration. Ashley and Abrams (1980) identified three mappable zones of alteration consisting of an intensely altered central silica cap surrounded by subsequently less altered zones of opalized and argillized rock (Figure 1). The mineralogy in the silicified zone was observed to be primarily quartz with minor calcite, alunite, and kaolinite. The opalized rocks contain the alteration minerals opal, alunite, and kaolinite. The argillized zone mineralogy consists of kaolinite derived from plagioclase, and montmorillonite and opal derived from volcanic glass.

ANALYSIS SOFTWARE

Complete analysis of imaging spectrometer data requires sophisticated image processing techniques that combine simultaneous extraction and display of both spectral and spatial information. Analysis software is under development at several locations that takes advantage of the combined high resolution graphics and imaging capabilities of modern image processing workstations (Mazer et al, 1988; Torson, 1989). An analysis package called "Integrated Software for Imaging Spectrometers" (ISIS) (Torson, 1989) was used for the analysis of the Cuprite GERIS and TM data. ISIS runs on a Digital Equipment Corporation (DEC) VAXstation with a GPX color graphics display configured with an additional 1024 x 1024 image display (IVAS, by International Imaging Systems). This combination allows simultaneous display of three spatial image planes as a color composite image on the IVAS display with a side slice showing the spectral dimension, and real-time display of spectra on the GPX graphics screen. A vertical line-cursor on the IVAS display shows the location of the spectral slice, while a standard crosshair-cursor shows the location of the current spectrum. Concurrent application processes allow selection of ground targets for calibration, extraction of average spectra, and spectral classification. CSES is developing concurrent ISIS analysis programs that utilize expert system capabilities (Kruse et al, 1988) and allow spectral unmixing to be used in the image classification (Boardman, 1989).

COMPARATIVE ANALYSIS OF TM AND GERIS DATA

Thematic Mapper

TM data have 30 meter spatial resolution and six broad spectral bands in the visible and near infrared. TM data also have a band in the thermal region which

was not used in this analysis. The scene used in this study was obtained in October 1984. The data were converted to exoatmospheric reflectance using look-up-tables published by EOSAT (Markham and Barker, 1986).

The TM data were analyzed using both traditional techniques and the ISIS software. False color composites, color ratio composites and traditional maximum likelihood classification techniques were used to analyze the data (Rowan et al, 1974; Lillesand and Kiefer, 1987). Both the classification and ratio images show the best results by allowing discrimination of general mineral groups. Unaltered and altered rocks are mapped as discrete units but individual mineral discrimination is not obtained. A classification was also performed on the TM data using the ISIS "spectrum ratioing" technique. This method classifies an image by dividing the spectrum for each pixel in the image by a reference spectrum extracted from the data. The resulting average deviation from 100% is compared to a tolerance to determine if the image spectrum matches the reference spectrum. This technique was used successfully to discriminate the iron bearing areas from non-iron areas in the TM data. This technique did not work well in discriminating between the alunitic and clay areas. Figure 2 shows the extracted spectra from the TM data for known mineralogical units. The TM spectra obviously undersample the available spectral detail and are very similar for the alunite, kaolinite, and buddingtonite. Only the hematite spectrum has spectral character that makes it different.

GERIS Data

The Geophysical and Environmental Research Imaging Spectrometer (GERIS) instrument is the first commercial imaging spectrometer. Imaging spectrometers measure light reflected from the Earth's surface, utilizing many narrow contiguous spectral bands to construct detailed reflectance spectra for millions of discrete picture elements (pixels) (Goetz et al, 1985). The GERIS collects data from 0.43 to 2.5 μm in 64 channels of varying width using three grating spectrometers with three individual linear detector arrays. The 24 visible and infrared bands between 0.43 and 0.972 μm are 23 nm wide, the 8 bands in the infrared between 1.08 and 1.8 μm are 120 nm wide, and the 31 bands from 1.99 to 2.5 μm are 16 nm wide (William Collins, written communication, 1988). The GERIS was flown over Cuprite during August 1987.

Analysis of the GERIS data in the spectral dimension results in extraction of absorption band information that allows definition of the surface mineralogy at Cuprite. Because complete spectra can be extracted from the imaging spectrometer data, it was possible to identify and map individual minerals using an ISIS "binary encoding" technique (Mazer et al, 1988). The GERIS data show a roughly concentrically zoned hydrothermal system (Figure 3). The mapped mineral zones do not correspond one-for-one to Abrams' alteration zones, however, a general match is observed. A central silica cap mapped previously by Abrams et al (1977) (Figure 1) was not distinguished using spectral classification of the 63 channel imaging spectrometer data because spectra of the cap were highly variable and no characteristic spectrum was identified. The imaging spectrometer data show that the first zone out from the silica cap consists primarily of alunite with absorption bands at 2.16, 2.32, and 2.42 μm (Figure 4). This corresponds with portions of the opalized zone mapped

by Abrams et al (1977). A kaolinitic zone is found farther from the center with an asymmetrical absorption band at 2.21 μm (Figure 4). This corresponds in part with the argillized zone of Abrams. Significant occurrences of alunite and kaolinite are also mapped west of U. S. Highway 95, however, these exposures are outside the area of mapped alteration and have not yet been evaluated in the field. Areas of ammonium enrichment near the northwest edge of the former hydrothermal system were identified by using the GERIS data to detect the presence of the mineral buddingtonite. Buddingtonite is an ammonium bearing feldspar discovered in the Cuprite district using the NASA Airborne Imaging Spectrometer (AIS) (Goetz and Srivastava, 1985). Although very difficult to recognize in the field, identification of buddingtonite was possible using the imaging spectrometer data because of the presence of a broad 2.117 μm absorption band and a secondary narrow band near 2.02 μm (Krohn, 1986; Krohn and Altaner, 1987) (Figure 4). Hematite mapped using the GERIS data (Figure 5) is distributed primarily around the contact between the altered and unaltered volcanic rocks.

SUMMARY

High spectral resolution remote sensing (imaging spectrometry) is a new tool that can be used to supplement existing geologic maps and to quickly produce detailed information for previously unmapped areas. Analysis of imaging spectrometer data allows extraction of a detailed spectrum for each picture element (pixel) of the image. Broad-band remote sensing systems, such as the Landsat Multispectral Scanner (MSS) and Landsat Thematic Mapper (TM), drastically undersample the information content available from a reflectance spectrum and identification of individual minerals is not possible. An imaging spectrometer, on the other hand, samples at close intervals and allows construction of spectra that closely resemble those measured with laboratory instruments (Figure 6). The high spatial and spectral resolution imaging spectrometer systems make identification of individual minerals and mineral assemblages possible.

Zoned mineralogy mapped at Cuprite, Nevada using the GERIS system conforms in a general sense to previous field and multispectral image mapping. Identification of individual minerals, however, using the imaging spectrometer data and spatial display of the dominant mineralogy adds information that can be used to help determine the morphology and genetic origin of the hydrothermal system. The imaging spectrometer data shows that the zoning at Cuprite has many characteristics similar to those described for large replacement-type alunite deposits (Hall and Bauer, 1983).

ACKNOWLEDGEMENTS

The authors would like to thank William Collins of Geophysical and Environmental Research Inc. for providing the Cuprite GERIS data for evaluation and the U. S. Geological Survey (Flagstaff) for providing analysis software. Development of imaging spectrometer analysis techniques was funded in part under NASA/JPL contract #958039. XRD characterization and measurement of reflectance standards was supported by NASA grant NAGW-1601.

REFERENCES CITED

1. Abrams, M. J., Ashley, R. P., Rowan, L. C., Goetz, A. F. H., and Kahle, A. B., 1977, Mapping of hydrothermal alteration in the Cuprite mining district, Nevada using aircraft scanner images for the spectral region 0.46-2.36 μm : *Geology*, v. 5, p. 713-718.
2. Ashley, R. P., and Abrams, M. J., 1980, Alteration mapping using multispectral images-Cuprite mining district, Esmeralda County, Nevada: *U. S. Geological Survey Open File Report* 80-367.
3. Boardman, J. W., 1989, Spectral and spatial unmixing: Applications of singular value decomposition: in *Proceedings, Image Processing '89*, 23-26 May, 1989, Sparks, Nevada (in press).
4. Goetz, A. F. H., and Srivastava, V., 1985, Mineralogical mapping in the Cuprite mining district: in *Proceedings of the Airborne Imaging Spectrometer (AIS) Data Analysis Workshop*, 68-10 April 1985, JPL Publication 85-41, Jet Propulsion Laboratory, Pasadena, California, p. 22-29.
5. Goetz, A. F. H., Vane, Gregg, Solomon, J. E., and Rock, B. N., 1985, Imaging spectrometry for earth remote sensing: *Science*, v. 228, p. 1147-1153.
6. Hall, R. B., and Bauer, C. W., 1983, Alunite: *Industrial Minerals and Rocks*, 5th edition, S. J. Lefond (ed.), AIME, p. 417-434.
7. Kahle, A. B., and Goetz, A. F. H., 1983, Mineralogical information from a new airborne thermal infrared multispectral scanner: *Science*, v. 222, no. 4619, p. 24-27.
8. Krohn, M. D., 1986, Spectral properties (.4 to 25 microns) of selected rocks associated with disseminated gold and silver deposits in Nevada and Idaho: *Journal of Geophysical Research*, v. 91B, no. 1, p. 767-783.
9. Krohn, M. D., and Altaner, S. P., 1987, Near-infrared detection of ammonium minerals: *Geophysics*, v. 52, no. 7, p. 924-930.
10. Kruse, F. A., Calvin, W. M., and Seznec, O., 1988, Automated extraction of absorption features from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Geophysical Environmental Research imaging spectrometer (GERIS) data: in *Proceedings AVIRIS performance evaluation workshop*, JPL publication 88-38, Jet Propulsion Laboratory, Pasadena, CA, p. 62-75.
11. Lillesand, T. M., and Kiefer, R. W., (eds.), 1987, *Remote Sensing and Image Interpretation*: Wiley & Sons, New York, p. 675-678.
12. Markham, B. L., and Barker, J. L., 1986, Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures: *Landsat Technical Notes*, No. 1, August, 1986, EOSAT, Lanham, MD., p. 3-8.
13. Mazer, A. S., Martin, M., Lee, M., and Solomon, J. E., 1988, Image processing software for imaging spectrometry data analysis: *Remote Sensing of Environment*, v. 24, no. 1, p. 201-210.
14. Rowan, L. C., Wetlaufer, P. H., Goetz, A. F. H., Billingsley, F. C., and Stewart, J. H., 1974, Discrimination of rock types and detection of hydrothermally altered areas in south-central Nevada by the use of computer enhanced ERTS images: *U. S. Geological Survey Professional Paper* 883, 35 p.
15. Shipman, Hugh, and Adams, J. B., 1987, Detectability of minerals on desert alluvial fans using reflectance spectra: *Journal of Geophysical Research*, v. 92, no. B10, p. 10391-10402.
16. Torson, J. M., 1989, Interactive image cube visualization and analysis: in *Proceedings, Chapel Hill Workshop on Volume Visualization*, 18-19 May, 1989, University of North Carolina at Chapel Hill, (in press).

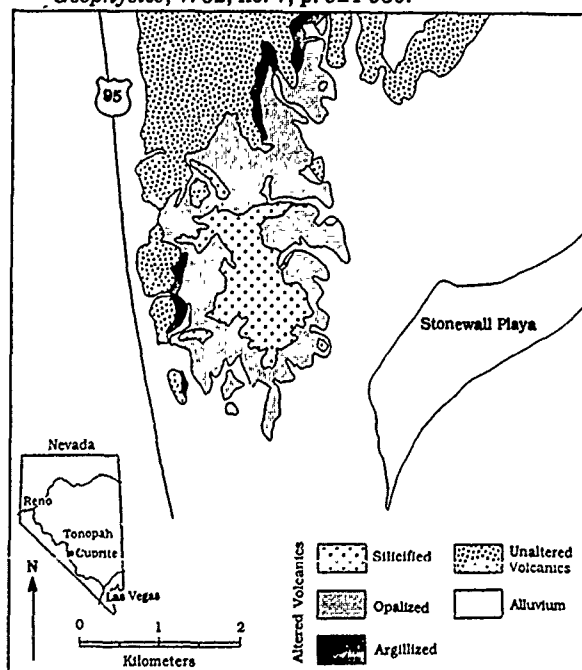


Figure 1. Alteration map for the Cuprite mining district. Redrawn from Shipman and Adams (1987) after Abrams et al (1977).

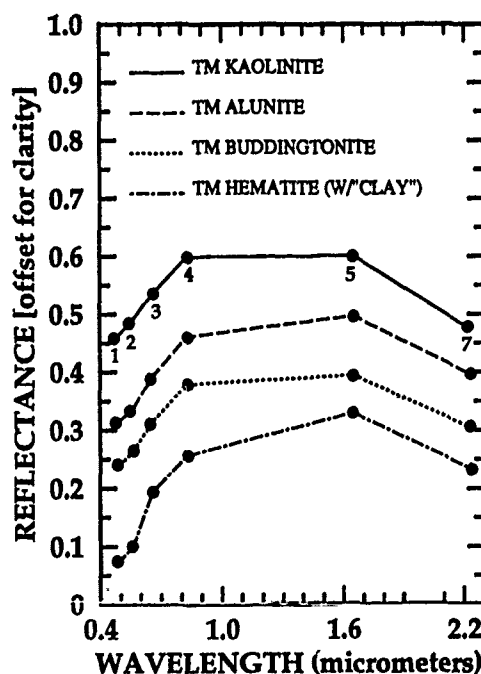


Figure 2. Cuprite, Nevada thematic mapper reflectance spectra for kaolinite, alunite, buddingtonite, and hematite. All four spectra are offset vertically for clarity.

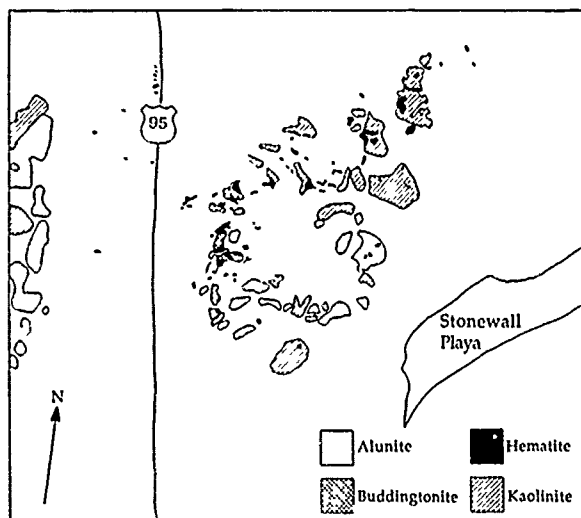


Figure 3. Mineral map derived from GERIS data. Note that the image was not geometrically corrected prior to interpretation and thus the orientation and geometry of the image does not exactly match that of the alteration map shown in figure 1. Comparison of the mineral map with the alteration map is facilitated by noting the locations and geometry of U.S. Highway 95 near the east edge of the image and Stonewall Playa near the center of the right edge.

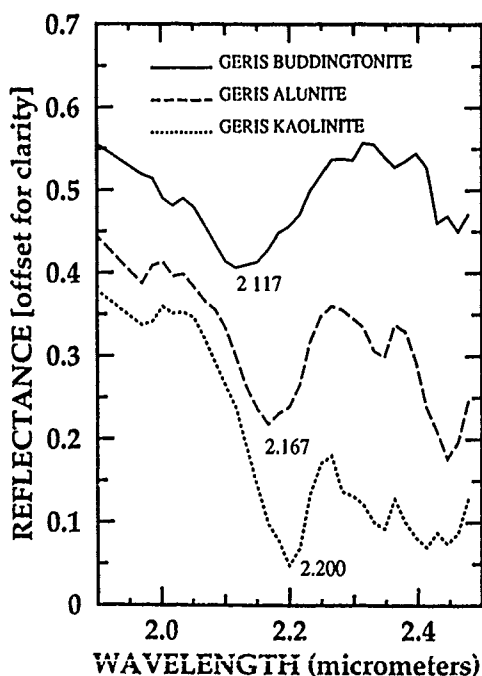


Figure 4. Comparison of GERIS buddingtonite, alunite, and kaolinite spectra showing the locations of the strongest absorption bands. All three spectra are offset vertically for clarity.

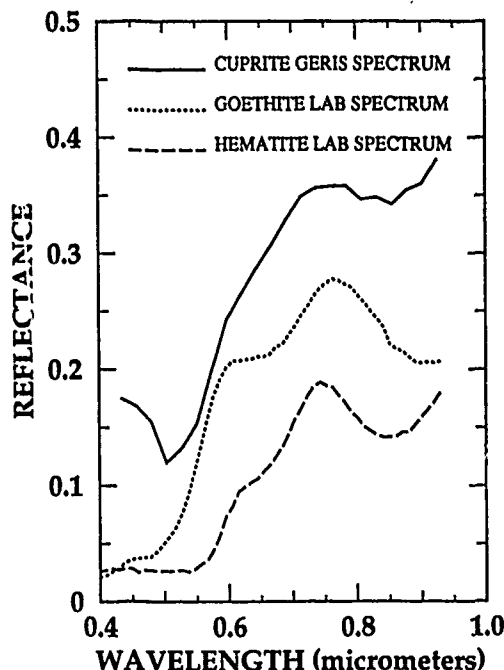


Figure 5. Comparison of laboratory spectra of hematite and goethite and a GERIS spectrum from Cuprite, Nevada interpreted to contain hematite. Note the match of the broad absorption feature near 0.85 μm . All three spectra are offset vertically for clarity.

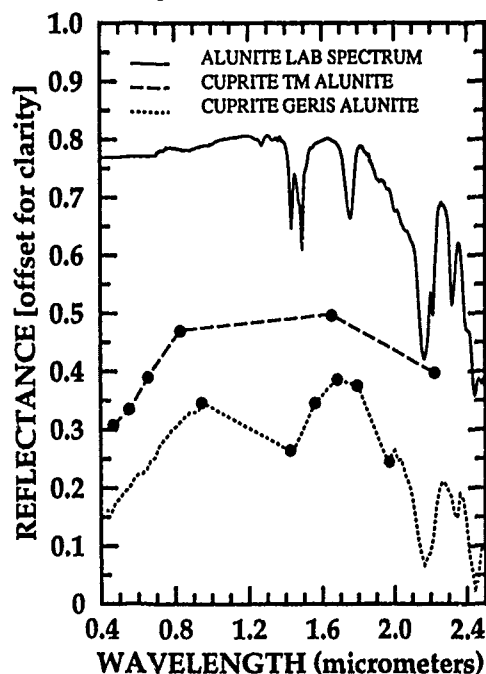


Figure 6. Comparison of a laboratory spectrum of alunite (Sulfur, California), a TM spectrum from Cuprite, Nevada, and a GERIS spectrum from Cuprite, Nevada. Note the loss of characteristic absorption band information in the TM spectrum. All three spectra are offset vertically for clarity.

IDENTIFICATION OF ILLITE POLYTYPE ZONING IN DISSEMINATED GOLD DEPOSITS USING REFLECTANCE SPECTROSCOPY AND X-RAY DIFFRACTION - POTENTIAL FOR MAPPING WITH IMAGING SPECTROMETERS

Fred A. Kruse and Phoebe L. Hauff
Center for the Study of Earth from Space (CSES)
CIRES, University of Colorado, Boulder, CO 80309
Phone (303) 492-6880, FAX (303)-492-5070

ABSTRACT

X-Ray diffraction studies of clay minerals from several disseminated gold deposits in Nevada demonstrate that illite polytypes are often laterally and vertically zoned around ore bodies. Polytypes act as geothermometers indicating temperatures of ore deposition and thus proximity to the hydrothermal fluids during deposition. The distribution of illite mineralogy was evaluated for several major disseminated gold deposits in Nevada using a field portable reflectance spectrometer. Changes in the visible/near-infrared reflectance spectra were observed as the illites progressed from the lower temperature, less ordered 1M variety through the higher temperature 2M type. Distinctive absorption features near 1.90 and 2.20 μm and between 2.3-2.5 μm in both field and laboratory spectra differentiated interlayer water, structural water (OH-), and some octahedral layer characteristics. Field reflectance spectrometry can be used to assist exploration efforts by providing detailed mineralogical information in real-time at the field location. As the new generation of imaging spectrometers is developed it is likely that subtle spectral differences such as those between the illite polytypes will become useful for remote exploration for mineral deposits.

INTRODUCTION

Reflectance spectroscopy and multispectral remote sensing are proven tools for locating and mapping hydrothermally altered rocks (Rowan et al, 1974; Marsh and Mckeon, 1983; Goetz et al, 1985; Lee, 1985; Hutinspiller, 1988; Feldman and Taranik, 1986; Kruse, 1988; Kruse et al, 1989). Broad band systems such as the Landsat Multispectral Scanner (MSS, 4 spectral bands) and Thematic Mapper (TM, 6 spectral bands) provide an operational capability that can be used by the exploration geologist to identify areas for more detailed study. However, these instruments identify only general mineral groups such as carbonates, clays, and iron oxides. New imaging spectrometer systems with higher spatial and spectral resolution (up to 224 narrow spectral bands) make identification of individual minerals (alunite, kaolinite, illite, hematite, goethite, etc.) and mineral assemblages possible thus permitting detailed site-specific mapping of alteration mineralogy.

The Carlin-Trend disseminated gold deposits of

north-central Nevada are ideal for study of alteration assemblages using field and laboratory spectral measurements and multispectral remote sensing because the geology is generally well known and gold assays already exist for many areas. The Carlin mine in Eureka County, Nevada, was selected for the initial study because it is recognized as the type locality for the original sediment hosted disseminated gold deposit. Acidic, gold-bearing hydrothermal solutions have extensively modified the host carbonates at Carlin. High grade ore is concentrated in weakly to moderately calcareous, weakly to moderately silicified Roberts Mountains Formation (Bakken and Enaudi, 1986).

X-RAY DIFFRACTION DETERMINATION OF ILLITE POLYTYPES

X-ray diffraction studies indicate that illite is an important mineral species in disseminated gold deposits, reflecting lateral and vertical zonation around gold ore bodies. Illites change crystal structure with temperature and may act as geothermometers indicating temperatures of ore deposition and proximity to hypogene and supergene fluids. Figure 1 is a compilation of X-ray diffraction spectra showing the two most common illite species or polytypes. These range from the higher temperature, well ordered 2M to the poorly ordered, lower temperature 1M. Illite polytypes can now easily be differentiated by automated X-ray diffraction. Several diagnostic polytype peaks occur between 23-32 degrees two-theta. These peaks and their hkl values are marked on figure 1. The 2M₂ from Japan (A) is a classic high temperature hydrothermal illite from Dr. S. Shimoda of the University of Tsukuba, Japan. The 2M illite from the Marblehead location (B) is of sedimentary origin and although identifiable as a 2M species, it is quite different from the Shimoda sample. There is considerably more water in its structure with interstratified smectite. This is shown by the broad, asymmetrical profile of its 10Å (8.9°) peak and the higher background from 2° - 7°. The Silver Hill sample (D), also sedimentary, has been identified as a 1M_d which indicates disorder in its stacking. The diagnostic 1M, smaller peaks at ~25.5° and ~27° are not as well defined as in the Fithian sample (C) which is considered to be 1M. It becomes obvious that there is extensive variation among the illites.

Preliminary work at Carlin indicates that illites vary

regularly with proximity to ore bodies. Figure 2 contrasts the X-ray diffraction data for a 2M illite (upper scan) from a high gold bearing zone with a 1M illite (lower scan) from a barren, silica flooded alteration zone at the Carlin Mine. The contrast between the polytypes is striking and shows potential as an exploration tool.

SPECTRAL MEASUREMENTS

Changes in the illite reflectance spectra are observed as the illites progress from the lower temperature, less ordered 1M variety through the higher temperature 2M type (Figure 3). Distinctive absorption features near 1.40, 1.90, and 2.20 μm and between 2.3-2.5 μm differentiate interlayer water, structural water (OH^-) and some octahedral layer characteristics. The reflectance spectrum for the 2M illite polytype is similar in appearance to that of muscovite (Hunt and Salisbury, 1970; Lee and Raines, 1984). The 2M illites typically have a sharp absorption feature at 2.2 μm caused by OH^- in the mineral structure (Hunt and Salisbury, 1970). In the 1M polytype, this band is broader and asymmetrical towards longer wavelengths. The 1Md spectrum from Silver Hill suggests that a secondary absorption feature near 2.25 μm may be the cause of the absorption band asymmetry. Additional weak absorption bands at 2.35 and 2.45 μm in illites also appear to be better developed in the 2M polytypes (Figure 3).

The presence of absorption bands at both 1.4 and 1.9 μm is indicative of molecular water as water of hydration or water trapped in the mineral lattice (Hunt and Salisbury, 1970). If the 1.4 and 1.9 μm bands are sharp, as they are in the 2M polytype (Figure 3), this indicates that the water molecules are located in well defined, well ordered sites. When these bands are broad and asymmetrical, as they are in the 1M illite polytypes, then they indicate that the mineral is relatively disordered (Hunt and Salisbury, 1970). The occurrence of the 1.4 μm band without the 1.9 μm absorption band indicates the presence of hydroxyl (OH^-) and the absence of molecular water.

The reflectance spectra for two illite polytypes taken with a portable reflectance spectrometer in the main Carlin pit are shown in Figure 4. Although the features are not as well defined as those observed on the laboratory spectra in Figure 3 the differences are apparent. Field spectrum C10885 is from alteration zone 3 (moderately silicified laminated beds) of Bakken and Einaudi (1986), a typically high gold zone. The well developed absorption bands at 2.2, 2.35, and 2.45 μm indicate that it is a 2M polytype. Field spectrum CAR88B is from alteration zone 5 (strongly silicified clastic beds) of Bakken and Einaudi (1986), a zone typically having low gold values. These spectra substantiate that the polytypes can be identified and mapped in the field and that there is potential for this method as an exploration and mapping tool.

CONCLUSIONS

The combined x-ray diffraction/reflectance spectroscopy study of illite polytypes has established that sufficient spectral differences exist to allow identification using near-infrared reflectance spectroscopy. Field and laboratory spectrometers have successfully been used to identify polytypes, however, the distinctions are quite subtle. Field reflectance spectroscopy can now be used operationally to assist exploration efforts by providing

detailed mineralogical information real-time at the field location. The potential exists, given sufficient spectral and spatial resolution and instrument signal-to-noise to identify these polytypes from the air with an imaging spectrometer.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the gracious cooperation and interest of the Newmont Gold Company for granting access to the Carlin property for the field study portion of this project and to thank Barbara Bakken for access to her sample suite. Analysis of the illite reference standards was funded in part by NASA grant NAGW-1601.

REFERENCES CITED

- Bakken, B. M., and Einaudi, M. T., 1986, Spatial and temporal relations between wall-rock alteration and gold mineralization, main pit, Carlin Gold Mine, Nevada: in *Proceedings "Gold 86"*, Vancouver, B.C., Canada, v. 1, Konsult International Inc, Willowdale, Ontario, Canada.
- Feldman, Sandra, and Taranik, J. V., 1986, Identification of hydrothermal alteration assemblages using Airborne Imaging Spectrometer data: in *Proceedings, 2nd Airborne Imaging Spectrometer (AIS) Data Analysis Workshop, 6-8 May, 1986, JPL Publication 86-35*, Jet Propulsion Laboratory, Pasadena, California, p. 96-101.
- Goetz, A. F. H., Vane, Gregg, Solomon, J. E., and Rock, B. N., 1985, Imaging spectrometry for earth remote sensing: *Science*, v. 228, p. 1147-1153.
- Hunt, G. R., and Salisbury, J. W., 1970, Visible and near-infrared spectra of minerals and rocks: I: Silicate Minerals, *Mod. Geol.*, v. 2, p. 283-300.
- Hutinspiller, Amy, 1988, Discrimination of hydrothermal alteration mineral assemblages at Virginia City, Nevada, using Airborne Imaging Spectrometer data: *Remote Sensing of Environment*, v. 24, no. 1, p. 53-66.
- Kruse, F. A., 1988, Use of Airborne Imaging Spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada and California: *Remote Sensing of Environment*, v. 24, no. 1, p. 31-51.
- Kruse, F. A., Kierein-Young, K. S., and Boardman, J. W., 1989, Mineral mapping at Cuprite, Nevada with a 63 channel imaging spectrometer: *Photogrammetric Engineering and Remote Sensing*, v. LV, no. 4, September, 1989 (submitted).
- Lee, Keenan, 1985, Interactive digital image analysis of Landsat images for mapping hydrothermal limonite: in *Proceedings, International Symposium on Remote Sensing of Environment, Fourth Thematic Conference, "Remote Sensing for Exploration Geology"*, Environmental Research Institute of Michigan, Ann Arbor, p. 293-307.

Lee, Keenan, and Raines, G. L., 1984, Reflectance spectra of some alteration minerals--a chart compiled from published data 0.4 μ m-2.5 μ m: *U.S. Geological Survey Open-File Report 84-96*, 6 p., 1 chart.

Marsh, S. E., and McKeon, J. B., 1983, Integrated analysis of high-resolution field and airborne spectroradiometer data for alteration mapping: *Econ. Geol.*, v. 78, no. 4, p. 618-632.

Rowan, L. C., Wetlaufer, P. H., Goetz, A. F. H., Billingsley, F. C., and Stewart, J. H., 1974, Discrimination of rock types and detection of hydrothermally altered areas in south-central Nevada by the use of computer enhanced ERTS images: *U. S. Geological Survey Professional Paper 883*, 35 p.

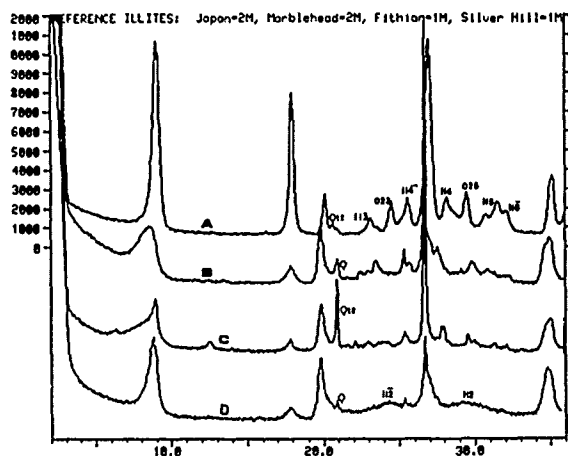


Figure 1. X-ray diffraction scans of reference illites showing polytype differences. Vertical scale shows intensities. Horizontal scale is in degrees two-theta. Peaks marked with *hkl* values indicate some diagnostic polytype reflections. A is a Japanese hydrothermal 2M; B a sedimentary 2M; C a sedimentary 1M; and D a disordered 1M.

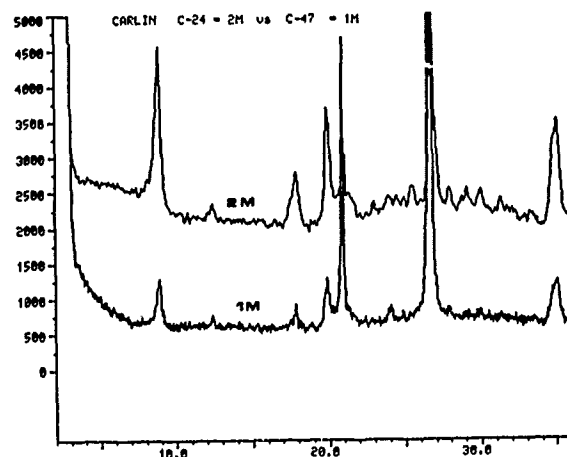


Figure 2. X-ray diffraction patterns for samples from the Main Carlin Pit. Vertical scale shows intensities. Horizontal scale is in degrees two-theta. Sample C-24, a well developed 2M illite, is from an ore zone and shows high gold values. Sample C-47 is a 1M disordered illite from a barren zone.

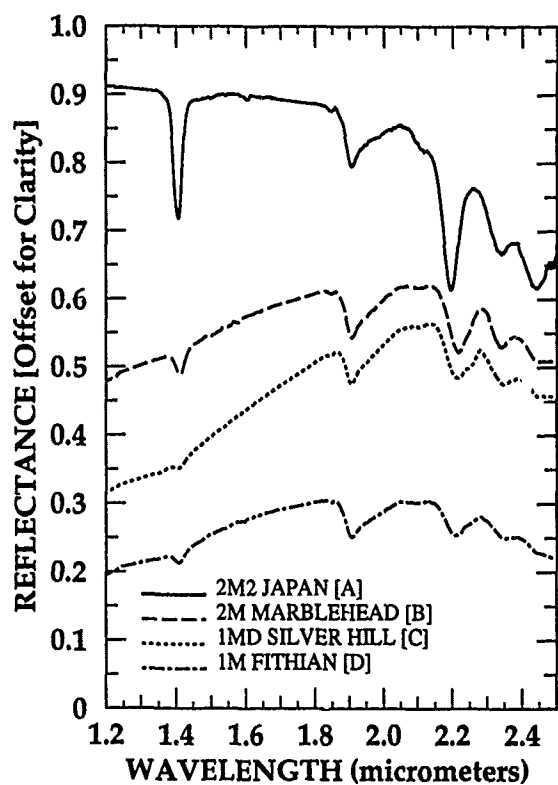


Figure 3. Reflectance spectra of illite reference standards showing polytype differences. Note how the absorption features at 1.4, 1.9, 2.2, and between 2.3 and 2.5 μm change with polytype. A is a higher temperature Japanese hydrothermal 2M; B a sedimentary 2M; C a sedimentary 1M; and D a disordered 1M. Spectra were collected on a Beckman 5270 laboratory spectrometer with integrating sphere.

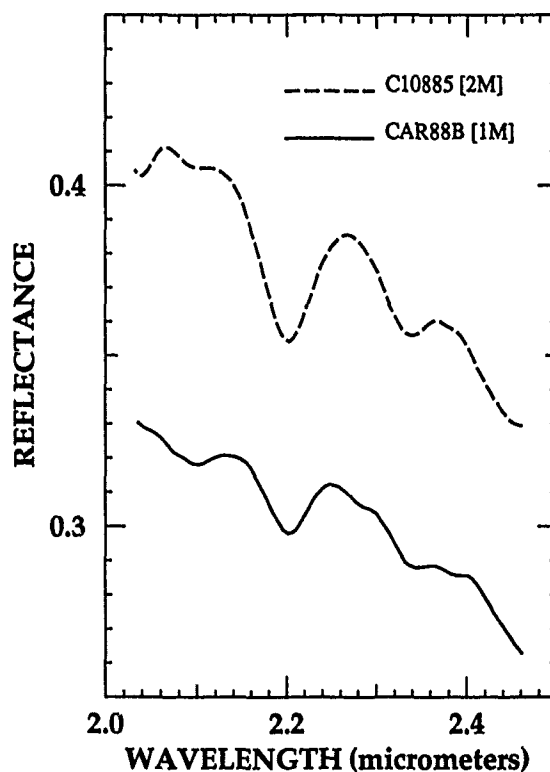


Figure 4. Illite reflectance spectra collected with a portable reflectance spectrometer in the Main Carlin Pit. The top spectra is a 2M illite while the bottom is a 1M. Note the differences in the definition of the features at 2.2 μm and from 2.3 μm to 2.5 μm .

Present Status and Operation History of MOS-1

Kohei Arai

Earth Observation Center,
National Space Development Agency of Japan
1401 Ohashi, Hatoyama, Saitama 350-03 Japan

ABSTRACT

The MOS-1 satellite was launched at 01:23(UT), Feb. 19 1987. After 3 months mission check, well qualified data of 3 sensors onboard MOS-1, MESSR, VTR, MSR and DCS transponder have been transmitted. Aside from mission instruments, a solar paddle generating power is in the specification while fuel still remains with a sufficient margin. The battery has been operating at the DOD of 13% in average with a little temperature deviation. Meanwhile MOS-1 orbit has been controlling within the range of 10km apart from nominal orbit except a few cases. Satellite attitude has been stabilizing within the range of the specification except a few cases. Major troubles on MOS-1 encountered so far are as follows. (1) The inversion in the stored command memory due to high energy particles including heavy ions. (2) Insufficient threshold level for the earth sensors.

1. Introduction

NASDA launched Marine Observation Satellite-1(MOS-1) on Feb. 19 1987 and get it into the expected orbit then in the 3 months mission check, confirmed the functions and performance of the mission and bus instruments on orbit. As for the initial phase of MOS-1, major specification and functions/performance were described in Ref.1 while for the results from assessment of mission instruments were also described in Ref.2 - 8. Since May 20 1987, MOS-1 has been operating routinely so far. It has been 2 years after the launch, design life of MOS-1.

First, MOS-1 operation will be summarized. Then detailed operation history such as orbital maneuvering, fuel consumption, degradation of battery, data transmission and acquisition, etc will be followed. Finally, major troubles encountered are described.

2. MOS-1 operation summary

According to the major purposes of MOS-1, MOS-1 operation are summarized as follows.

(1) to establish fundamental technologies on earth observation satellites

Except one case of satellite attitude loss occurred on July 28 1987, functions and performance of mission and bus instruments were confirmed within the range of the specification.

(2) to observe mainly ocean phenomena using the mission instruments

Usefulness of mission instruments data has been confirming in the MOS-1 verification program.

(3) to conduct primary experiments with DCS transponder

Data link success rate and determination accuracy of DCP location were confirmed. For the DCP with output power is higher than 4 W, the success rate is higher than 80% while the determination accuracy is less than 500m if the DCPs are located within the range from 100 - 2500km against from nominal orbit.

(4) to establish technologies on tracking and control of the satellite in a sun synchronous orbit

In order to maintain the designated sun synchronous recurrent orbit, satellite altitude and inclination have been controlled. Thus WRS and LMT have been maintained within the range of the specification. Meanwhile orbit determination has also been performing with the specified accuracy.

(5) to establish technologies on operation of an earth observation satellite

10 ground stations in the world for

MOS-1 data acquisition have been operating. Mission management including mission instruments operation planning has been doing well.

3. Detailed Operation History

3.1 Satellite system

3.1.1 Orbit control

Referred to the section 3.2.

3.1.2 Attitude control

The specified and evaluated are shown in Table 1.

average and no degradation on output voltage.

3.1.7 Attitude and orbit control subsystem

(1) Wheel unloading confirmed

(2) Orbit control

Satellite altitude control, inclination control functions were confirmed in 15 times. In particular, when the satellite got into the safe hold mode in July 1987, both aforementioned functions were confirmed.

Table 1 The specified and evaluated attitude error and attitude change rate

| The specified attitude error | The evaluated attitude error |
|------------------------------------|------------------------------------|
| 0.6 deg. for roll | 0.25 deg. for roll |
| 0.6 deg. for pitch | 0.33 deg. for pitch |
| 1.0 deg. for yaw | 0.69 deg. for yaw |
| The specified attitude change rate | The evaluated attitude change rate |
| 0.02 deg./sec for roll | 0.014 deg./sec for roll |
| 0.02 deg./sec for pitch | 0.004 deg./sec for pitch |
| 0.05 deg./sec for yaw | 0.006 deg./sec for yaw |

3.1.3 Fuel consumption

MOS-1 used about half of fuel so that approximately 9kg of fuel still remains. It is enough fuel for another 2 years hopefully.

3.1.4 Mission instruments operation

Three radiometers, MESSR, VTIR and MSR simultaneous operation for 15min./revolution, two radiometers simultaneous operation, single radiometer operation and two MESSR system simultaneous operation have been confirmed. For eclips, either VTIR or MSR operation has been confirmed. On the other hand, DCS transponder has been operating always.

3.1.5 TTC subsystem

S-band real time telemetry, playback telemetry, ranging signal MSR data were well qualified. VHF real time telemetry, playback telemetry and beacon signal were also confirmed within the range in the specification. Also the commanding functions were confirmed except single event upset for stored command memory.

3.1.6 Power supply

Stability of bus voltage were specified the range from 28.4 - 29.6V while the results showed 28.9V for eclips, and 29.3V in sun litte. As for the battery, 25% of DOD in maximum and about 10,000 cycle of charge and discharge are specification. Results shows approximately 13% of DOD in

3.1.8 Solar array paddle

The specification of generation power is more than 640W(BOL) and more than 540W(EOL) while for two years design life, more than 700W of power has been generating.

3.1.9 Thermal control

Through careful check of thermal telemetry data, all the thermal control functions were confirmed.

3.2 Orbit control operation

Fig.1 shows ground track deviation from the nominal orbit designated as WRS. MOS-1 orbit has been maintaining within the range of +/-10 km from the nominal orbit except 2 cases, with 4 times of satellite altitude maneuver. MOS-1 attitude was lost in July 1987 due to the out of range of measurable temperature of the Earth Sensor Assembly(ESA). Beside this MOS-1 orbit was out of the range of +/-10 km when the inclination of the satellite orbit was controlled in April 1988.

Local Mean Time(LMT) of MOS-1 orbit has been maintaining within the range from 10:00 - 10:30. Fig.2 shows LMT changes for the past and the future.

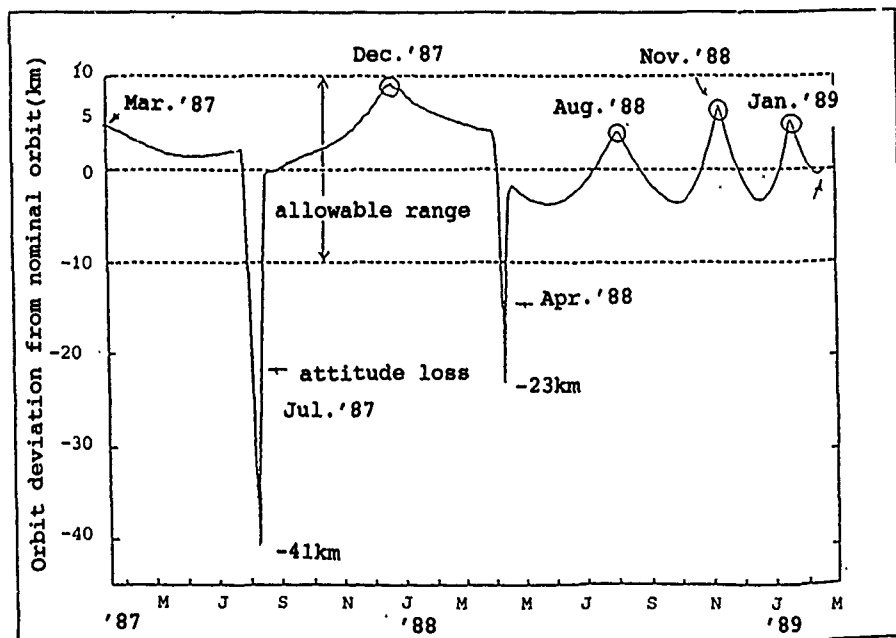


Fig.1 MOS-1 orbit control history

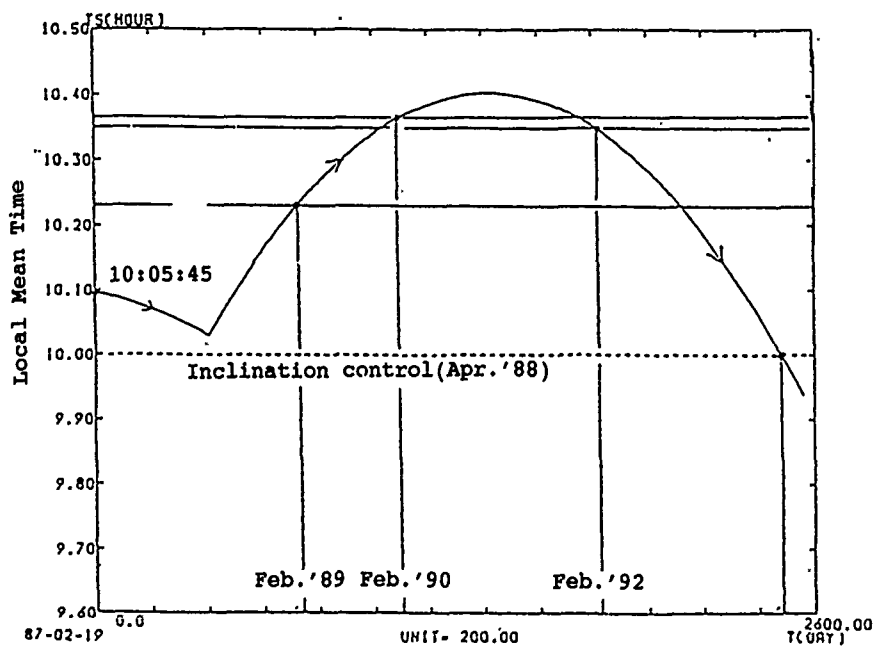


Fig.2 LMT changes

3.3 Major troubles encountered

Major troubles encountered so far are as follows.

(1) Single Event Upset (SEU)

Due to the aforementioned reason, SEU has been occurring approximately once a day. To avoid this, the stored command memory has been using outside of south atlantic anomaly area.

The stored command memory to be onboard MOS-1b is modified to avoid it.

(2) Malfunction on earth presence check function

MOS-1 got into the safe hold mode (Sun Acquisition Mode : SAM) through the detection of abnormal attitude error signal from Attitude and Orbit Control Subsystem : AOCS, on July 28 1987. To avoid this, MOS-1 has been operating without the earth presence check, for the period from June to September when the abnormal attitude error would occurred.

Threshold level of the earth sensor for MOS-1b is to be changed to avoid this.

4. Summary report from sensor validations

4.1 MESSR

4.1.1 Geometric accuracy

(1) Band to band registration

Using 10 - 20 GCPs/scene, band to band registration was evaluated with several conditions on the temperature at the MESSR optics and CCD, for more than 20 scenes. The results are shown in Table 2.

(2) Geometric fidelity

The aforementioned experiments have been conducted to evaluate geometric accuracy. The results shows about 4 - 5km in terms of rms error exist in MESSR image in average. Through a careful analysis on error vectors, bias components in roll, pitch and yaw directions were confirmed. After the correction of bias errors, it is found that 0.4km of rms residual error can be achieved in average.

4.1.2 Radiometric accuracy

It was confirmed that radiometric accuracy of MESSR was within one quantization bit. Through a careful analysis with calibration data (MESSR data in eclips), dark current component was also evaluated. Using MESSR data of homogeneous areas, radiometric accuracy was evaluated by normalizing the standard deviation with the average of MESSR data in each homogeneous area.

4.2 VTIR

4.2.1 Geometric accuracy

(1) Band to band registration

By means of the aforementioned procedure, band to band registration was evaluated. Table 3 shows the results.

(2) Geometric fidelity

Similar geometric accuracy in terms of rms error of VTIR data is approximately 6km for more than 20 scenes. After the correction of bias error, 2.5km of geometric accuracy is achieved.

4.2.2 Radiometric accuracy

Using similar manner to MESSR, one quantization bit of radiometric accuracy was confirmed for more than 20 scenes.

4.3 MSR

4.3.1 Geometric accuracy

23km of geometric accuracy of MSR data was found.

4.3.2 Radiometric accuracy

It was found that radiometric accuracy of MSR data was within 2K.

5. MOS-1 operation

MOS-1 has been operating for 10 ground stations (G/S), 2 G/S in Japan, 4 G/S in ESA, 1 G/S in Thailand, Canada, Australia, and Antarctica as is shown in Fig.3.

Table 2 Band to band registration of MESSR

| Direction | Line | Pixel | Remarks |
|-----------|------------|------------|-----------------------------|
| | 0.46 pixel | 0.21 pixel | in maximum for all the band |

Table 3 Band to band registration of VTIR

| Direction | Line | Pixel | Remarks |
|-----------|------------|------------|---------------------------------------|
| | 1.21 pixel | 0.54 pixel | By comparing band 3 and 4 with band 1 |

6. Concluding remarks

MOS-1 has been operating with sufficient remaining fuels, generating power, battery conditions, mission instruments performance, and so on. It is expected a longer mission life for MOS-1, than expected design life and also expected simultaneous operation with MOS-1b to be launched in the early 1990. Mission parameters for MOS-1b are same as MOS-1 except modification of threshold level of earth sensor assembly, stored command memory, minor modification of command codes, and orbit. MOS-1b will be put into the orbit of approximately 180deg. phase of MOS-1 orbit.

Reference

(1) Arai K and C.Ishida : On a verification plan for MOS-1, The 8th Canadian Symposium on Remote Sensing, Session 9, No.991, 1983.

(2) Arai K, C.Ishida and H.Wakabayashi : MOS-1 verification experiment, ISPRS Working Group II-3,4 Joint Meeting, B-18, 1983.

(3) Arai K, T.Igarashi and C.Ishida : Some results on field experiments in MOS-1 verification program, The 14th ISTS, m-1-1, 1984.

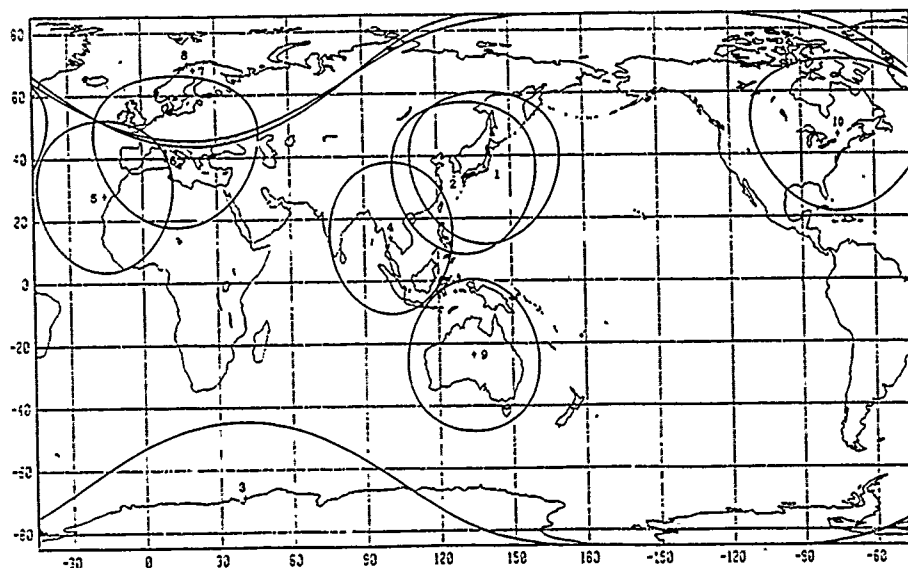
(4) Arai K, T.Igarashi, C.Ishida, H.Wakabayashi and Y.Takagi : Evaluation of MOS-1 MSR data in field experiments, The 18th ERIM Symposium, 1984.

(5) Tsuchiya K, K.Arai and T.Igarashi : Marine Observation Satellite - 1 (MOS-1), "Remote Sensing Reviews", Gordon Breach Science Publish Co., Ltd, 1985.

(6) Arai K., T.Igarashi and Y.Takagi : An emissivity model of snowpack for passive microwave observations, The 36th IAP, 85 - 98, 1985.

(7) Arai K : A preliminary assessment of radiometric accuracies for MOS-1 sensors, International J. Remote Sensing, vol.9, no.1, 5 - 21 1987.

(8) Arai K : MOS-1 system overview, The 17th LGSOWG Meeting, 1987.



| organization | location | MESSR/VTIR/MSR | | | start | status |
|----------------|-----------------------------|----------------|---|---|---------------|---------------------------|
| 1. EOC, NASDA | Hatoyama, Japan | X | X | X | Feb. 23, 1987 | operational |
| 2. Tokai Univ. | Kumamoto, Japan | X | X | | May 31, 1987 | operational |
| 3. NIPR | Showa base Antarctica | X | X | X | Feb. 1989 | scheduled |
| 4. NRCT | Bangkok, Thailand | X | X | X | Jul. 1, 1988 | operational |
| 5. ESA | Maspalomas, Spain | X | X | X | Nov. 2, 1987 | operational |
| 6. ESA | Fucino, Italy | X | X | X | Nov. 25, 1987 | operational |
| 7. ESA | Kiruna, Sweden | X | X | X | Nov. 2, 1987 | operational |
| 8. ESA | Tromso, Norway | | | X | Nov. 2, 1987 | operational |
| 9. CSIRO | Alice Springs, Australia | X | X | X | Jul. 1, 1988 | operational (6 months) |
| 10. CCRS | Gatineau, Canada | X | X | X | May 16, 1988 | operational (5 months) |

Fig.3 MOS-1 Ground Stations

Interband Correlations of Visible and Thermal Infrared Radiometer Imagery from the Marine Observation Satellite-I*

Matthew Heric

Autometric, Incorporated
5301 Shawnee Road
Alexandria, Virginia, 22312-2312 USA

ABSTRACT

Using relatively broad band remote sensing systems surface temperature phenomena can be difficult to discern. In particular, because of the often specific nature of thermal features, interband correlation levels of resultant data sets may limit the utility of multispectral infrared imagery. In response to this concern, Visible and Thermal Infrared Radiometer imagery from the Japanese Marine Observation Satellite-I was analyzed for amounts of interband correlation. Correlation coefficients were calculated and equal size density sliced single-band images were generated.

KEY WORDS: Correlation, VTIR, Thermal, and Density Slice.

INTRODUCTION

The Japanese Marine Observation Satellite (MOS-I) was successfully launched by a NIPPON N-II rocket from the Japanese Space Center at Tanegashima on 19 February 1987. MOS-I is equipped with three sensors, and among these is the Visible and Thermal Infrared Radiometer (VTIR). The VTIR sensor is designed to detect snow and ice distributions, atmospheric phenomena, and most notably, sea surface temperatures.

Thermal energy emitted from sea water can be difficult to discern. Indeed, in some instances sea surface temperature variations are so slight that broad band thermal systems may not provide data with sufficient accuracies to allow definite thermal differentiations. Systems with multiple thermal bands, such as the VTIR, can be used to clarify these discrepancies; but again, because thermal information is often so specific, broad band systems may experience significant interband correlation problems.

BACKGROUND

Designed for marine-related investigations, the introduction of MOS-I reflects the growing global interest in civil remote sensing. The VTIR, similar to the United States' Advanced Very High Resolution Radiometer (AVHRR/2), is designed to image the

earth using one band in the visible and three bands in the infrared portions of the electromagnetic spectrum. VTIR and AVHRR/2 system parameter comparisons are provided in Table 1.

Table 1

Comparisons of Principal Visible and Thermal Infrared Radiometer and Advanced Very High Resolution Radiometer/2 System Parameters

| | | Spectral | | Spatial | |
|-----------|---|----------|---|---------|--------------|
| VTIR* | | | | | |
| Band 1 | - | 0.50 | - | 0.70 | 0.9 x 0.9 km |
| Band 2 | - | 6.00 | - | 7.00 | 2.7 x 2.7 km |
| Band 3 | - | 10.50 | - | 11.50 | 2.7 x 2.7 km |
| Band 4 | - | 11.50 | - | 12.50 | 2.7 x 2.7 km |
| AVHRR/2** | | | | | |
| Band 1 | - | 0.58 | - | 0.68 | 1.1 x 1.1 km |
| Band 2 | - | 0.725 | - | 1.10 | 1.1 x 1.1 km |
| Band 3 | - | 3.55 | - | 3.93 | 1.1 x 1.1 km |
| Band 4 | - | 10.50 | - | 11.30 | 1.1 x 1.1 km |
| Band 5 | - | 11.50 | - | 12.50 | 1.1 x 1.1 km |

(Sources: *Koizumi, 1987; **Jensen, 1986)

* The MOS-I data used herein were produced by the National Space Development Agency of Japan.

One of the more useful applications of VTIR imagery is the detection of sea surface temperatures. As previous studies have shown, however, (Shaw and Irbe, 1972; Chahine, et al., 1983; and Estes, et al., 1983) surface temperatures are not obtained directly from thermal imagery. Rather, these distributions are derived by performing a number of corrections to the original data due to influences of the propagation path. While numerous algorithms have been written to accomplish these corrective procedures (Walton, 1987), energy emitted from sea water can remain difficult to analyze accurately using relatively broad band systems. Due to the engineering parameters of these systems, the sensitivities of the broad bands are not always great enough to discriminate specific thermal phenomena and, in turn, often generalize sensed levels of emitted energy. With multispectral imagery, the effects of this generalization can be reflected by the interband correlation levels of the data.

ANALYSIS TECHNIQUES

Noting potential difficulties in using data from broad band radiometers, VTIR imagery of Japan and the surrounding ocean areas was analyzed for interband correlation levels. The VTIR imagery used for this purpose was obtained by MOS-I on 1 June 1987. The correlation, or interrelation, levels of primary interest were for bands 2, 3, and 4. Band 1 (0.50 to 0.70 micrometers), generally sensitive to the visible portion of the electromagnetic spectrum, was included in order to keep the VTIR data set unbroken.

To calculate the interrelation of these bands, the correlation coefficient equation (Cooley and Lohnes, 1971) was used to ratio the covariance of any two VTIR bands to the product of their standard deviations as follows:

$$r(a)(b) = \frac{\text{COV}(a)(b)}{\text{SD}(a) * \text{SD}(b)},$$

where:

r = correlation coefficient;

$\text{COV}(a)(b)$ = covariance of bands a and b;

$\text{SD}(a)$ = standard deviation of band a;

and

$\text{SD}(b)$ = standard deviation of band b.

Therefore, this ratio was used to determine the sensitivities and separation abilities of each VTIR broad band.

While the correlation coefficient equation helped measure levels of interrelation, an additional method involved density slicing the respective single-band images. While alone this method was not considered a sufficient representation of correlation, it nevertheless provided a useful visual por-

trayal of the interrelations. Density slicing is the conversion of the continuous tones of an image into a series of discrete class intervals or slices where each interval corresponds to a specific brightness value range (Haralick, 1973; Jensen, 1986). This density slicing was accomplished based on the minimum and maximum unenhanced 8-bit brightness values of the VTIR original data. Therefore, while 256 gradients were available for this slicing, the absolute minimum and maximum values of the imagery were 1 and 150. In turn, 10 classes were formed by the equal range method as shown in Table 2.

Table 2

Intervals Used for Class Creation and Density Slicing Based on the Unenhanced Minimum and Maximum Brightness Range of the VTIR Data

| Original Digital Counts (input) | Class Number | Density Sliced Digital Counts (output) |
|---------------------------------|--------------|--|
| 1 - 15 | 1 | 40 |
| 16 - 30 | 2 | 60 |
| 31 - 45 | 3 | 80 |
| 46 - 60 | 4 | 100 |
| 61 - 75 | 5 | 120 |
| 76 - 90 | 6 | 140 |
| 91 - 105 | 7 | 160 |
| 106 - 120 | 8 | 180 |
| 121 - 135 | 9 | 200 |
| 136 - 150 | 10 | 220 |

As described previously, thermal measurements are not obtained directly from VTIR imagery but rather from a series of equations which remove atmospheric influences. Weinreb and Hill (1980), Walton (1987), and Maturi, et al. (1986) provided methods for radiometrically correcting thermal imagery for atmospheric attenuation; however for this VTIR investigation the computed absorption coefficients for water vapor and atmospheric gas removal were considered to be inaccurate due to inadequate regional weather temperature, pressure, and vapor content ground truth data. For these reasons, the resulting density sliced data could not be considered as absolute thermal readings. Instead, the classes represented digital counts and not radiance values.

RESULTS

The correlation coefficients of the VTIR data are provided in Table 3. Band 1 was not correlated significantly with bands 2, 3 and/or 4. As expected, this low level of correlation for band 1 suggested substantial differences in the visible and infrared data provided by the VTIR. Similarly, the correlations between bands 2 - 3 and 2 - 4 were also low. Band 2 was designed primarily for atmospheric investigations and not surface

thermal studies, and this, predictably, had a direct influence on these low correlations. By comparison, however, bands 3 and 4 were significantly correlated. Indeed, at 0.9709, this measure suggested that as a multispectral thermal infrared data set, the potential utility of these bands would be limited. While certainly valuable data were present in bands 3 and 4, their use would be restricted greatly.

Table 3

Calculated Correlation Coefficients (r)
for the VTIR data set

| Bands | 1 | 2 | 3 | 4 |
|-------|---|-------|--------|--------|
| 1 | X | .3006 | -.3718 | -.3509 |
| 2 | | X | .1307 | .1921 |
| 3 | | | X | .9709 |
| 4 | | | | X |

In addition to the correlation coefficients, Table 4 and Figure 1 show the minimum-to-maximum distributions for each band and accompanying density sliced images. The resultant sliced images echoed the calculated correlations. Again, the correlation levels of bands 3 and 4 were substantial. By comparison, bands 1 and 2 proved to be distinctive. The visual appearance of these density sliced images was beneficial to the understanding of the interband correlations.

Table 4

Resultant VTIR Density Sliced
Class Percentages

| Classes | Band 1 | Band 2 | Band 3 | Band 4 |
|---------|---------|---------|---------|---------|
| 1 | 37.41% | 18.67% | 0.17% | 0.23% |
| 2 | 36.18 | 52.82 | 0.30 | 0.34 |
| 3 | 9.25 | 28.50 | 0.63 | 0.65 |
| 4 | 7.39 | 0.01 | 0.95 | 0.64 |
| 5 | 6.01 | 0.00 | 1.39 | 2.19 |
| 6 | 2.51 | 0.00 | 5.33 | 5.43 |
| 7 | 0.76 | 0.00 | 31.02 | 34.83 |
| 8 | 0.36 | 0.00 | 41.84 | 43.40 |
| 9 | 0.10 | 0.00 | 16.46 | 11.82 |
| 10 | 0.03 | 0.00 | 1.91 | 0.47 |
| TOTALS | 100.00% | 100.00% | 100.00% | 100.00% |

CONCLUSIONS

By calculating the correlation coefficients the levels of interrelation for this specific VTIR imagery set were determined. The results indicated that as a 4-band multispectral data set, the high level of correlation

between bands 3 and 4 limited the potential utility of the multispectral imagery. The difficulty in discerning thermal energy using broadband systems was suggested by this relatively high correlation level.

REFERENCES

Chahine, M., et al., Manual of Remote Sensing. Ed. D. Simonett. Falls Church, VA: American Society of Photogrammetry and Remote Sensing. Vol. 1, pp. 165 - 23, 1983.

Cooley, W., and P. Lohnes, Multivariate Data Analysis. New York: John Wiley and Sons, Inc., pp. 49-97, 1971.

Estes, J., E. Hajic, and L. Tinney, Manual of Remote Sensing. Ed. D. Simonett. Falls Church, VA: American Society of Photogrammetry and Remote Sensing. Vol. 1, pp. 987 - 1124, 1983.

Haralick, R., "Glossary and Index to Remotely Sensed Image Pattern Recognition Concepts," Pattern Recognition, Vol. 5, pp. 391 - 403, 1973.

Jensen, J., Introductory Digital Image Processing. Englewood Cliffs, New Jersey: Prentice Hall, p. 21, 1986.

Koizumi, S., ed. MOS-I Data User's Handbook. Tokyo: Earth Observation Center - National Space Development Agency of Japan, pp. 2-12, 1987.

Maturi, E., J. Pritchard, and P. Clemente-Colon, An Experimental Technique for Producing Moisture Corrected Imagery from 1 Km Advanced Very High Resolution Radiometer (AVHRR) Data.

NOAA Tech. Report NESDIS 15, U.S. Department of Commerce, Washington, D.C., 1986.

Shaw, R., and G. Irbe, "Environmental Adjustments for the Airborne Radiation Thermometer Measuring Water Surface Temperature," Water Resources Research, Vol. 8, No. 5, pp. 1214 - 1225, 1972.

Walton, C., The AVHRR/HIRS Operational Method for Satellite Based Sea Surface Temperature Determination. NOAA Tech. Report NESDIS 28, U.S. Department of Commerce, Washington, D.C., 1987.

Weinreb, M., and M. Hill, Calculation of Atmospheric Radiances and Brightness Temperatures in Infrared Window Channels of Satellite Radiometers. NOAA Tech. Report NESDIS 80, U.S. Department of Commerce, Washington, D.C., 1980.

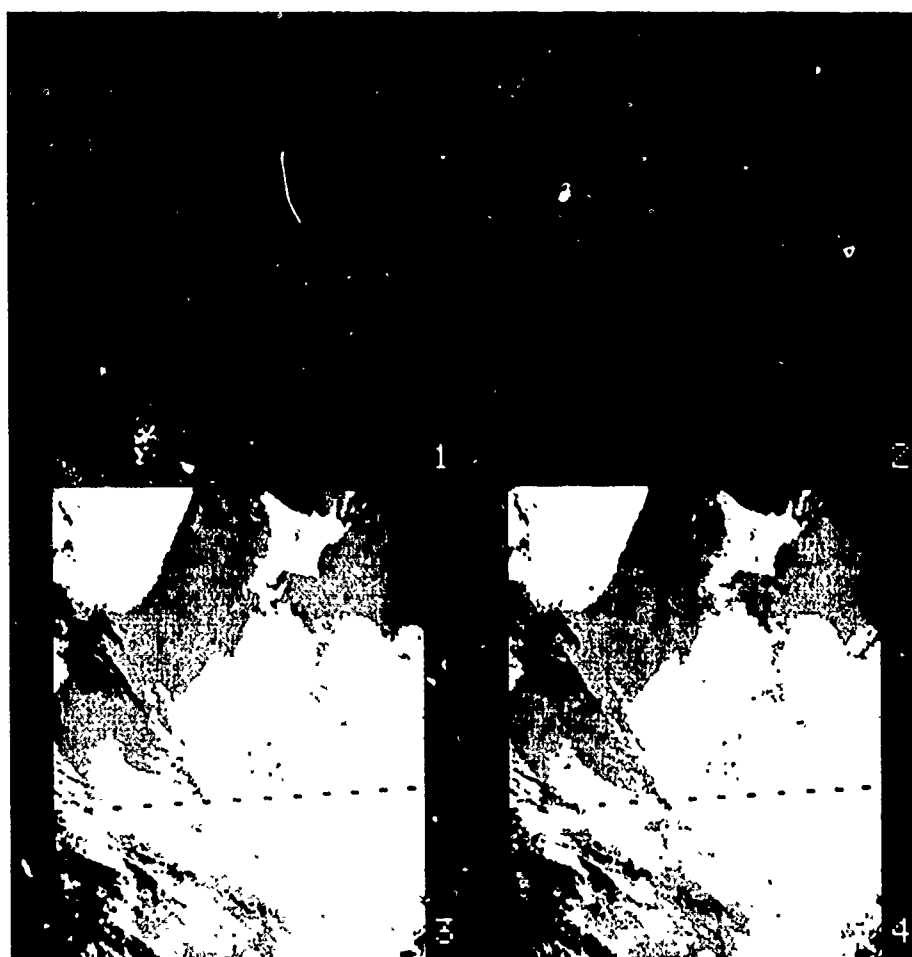


Figure 1. Density Sliced VTIR Single-Band Images.

VERIFICATION RESULTS OF MOS-1 MICROWAVE SCANNING RADIOMETER (MSR)

Korehiro Maeda, Michio Notomo, Shinji Ogawa
and Hideo Sato

Earth Observation Center (EOC),
National Space Development Agency of Japan
(NASDA)

ABSTRACT

Marine Observation Satellite-1 (MOS-1) was launched by N-II launch vehicle from Tanegashima Space Center, NASDA on Feb.19, 1987 and two years (design life of MOS-1) have passed. MOS-1 is still in satisfactory condition and normally operated. It is expected that MOS-1 will be operated more than one year.

MOS-1 has 4 mission instruments (MESSR, VTIR, MSR and DCST). The MSR is microwave scanning radiometer which is a Dicke type radiometer to measure very weak earth radiation noise at the 23 GHz and 31 GHz. As a part of MOS-1 Verification Program (MVP), NASDA evaluated geometric and radiometric performance of MOS-1 MSR by using truth data and airborne MSR data.

In this paper, outline of verification results of MOS-1 MSR is presented.

Keywords: MOS-1, Microwave Scanning Radiometer, MSR, MOS-1 Verification Program (MVP)

1. INTRODUCTION

Marine Observation Satellite-1 (MOS-1) is Japanese first observation satellite which was developed by using Japanese own technologies and successfully launched by N-II launch vehicle from Tanegashima Space Center, NASDA at 10:23 on Feb.19, 1987. MOS-1 has 4 mission instruments (MESSR: Multi-spectral Electronic Self Scanning Radiometer, VTIR: Visible and Thermal Infra-red Radiometer, MSR: Microwave Scanning Radiometer and DCST: Data Collection System Transponder). In order to evaluate MOS-1 observation system, NASDA has conducted MOS-1 Verification Program (MVP) in collaboration with joint research organizations and MVP participating organizations since March, 1987.

As a part of MVP, NASDA conducted MVP airborne experiments on MOS-1 passing days. NASDA evaluated MOS-1 sensors by using truth data, airborne data and others.

In this paper, outline of verification results of MOS-1 MSR is presented. Geometric and radiometric performances of MSR and capability to observe water vapor, cloud liquid, snow and ice distribution and others are clarified.

2. VERIFICATION FOR INSTRUMENTATION

2.1 Outline of MSR

The MSR is a Dicke type radiometer to measure very weak earth radiation noise at the 23 GHz and 31 GHz. Swath width is 317 km and beam widths are 32 km (23 GHz) and 23 km (31 GHz). Integration time for MSR data is 10 msec and 47 msec. The offset Cassegrain antenna used for both frequencies is conically rotated at about 18.75 rpm (scan period 3.2 sec) and observation is made for half of the antenna rotation (back side) and calibration signal such as high and low temperatures is inserted for the half of rotation (forward side). Table 1 shows performance of MOS-1 MSR and airborne MSR.

Table 1. Performance of MSR

| Item | Performance | |
|----------------------|--|---------------------------|
| Frequency | 23.8 ± 0.2 GHz | 31.4 ± 0.25 GHz |
| Antenna type | Offset Cassegrain | |
| Polarization | Horizontal (Vertical)* | Vertical (Horizontal)* |
| Beam width | 1.89° | 1.31° |
| Offset angle | 10°(30°) | 10°(30°) |
| Diameter of antenna | 50cm | |
| Scanning Period | 3.2 sec | |
| Angle of cone | 20° | |
| Swath width | 317 km (Altitude 909km) 2441 m (Altitude 7km) | |
| Receiver type | Dicke | Dicke |
| Receiver sensitivity | <1 K at 300K | <1 K at 300K |
| Dynamic range | 30 - 330 K | |
| Quantization level | 1024 (10 bits) | |
| Integration time | 10 & 47 msec | 10 & 47 msec |

* Addition of function for airborne MSR

2.2 Radiometric performance

(1) Radiometric distortion evaluation

Mean and standard deviation in Ocean area of MSR imagery (level 2) were measured. Example of standard deviation measured were 1.14K (23GHz 10msec), 1.13K (23GHz 47msec), 1.33K (31GHz 10msec), 1.30K (31GHz

47msec). In case of evaluation of internal calibration source, the deviations were 0.579 K (23GHz) and 0.616 K (31GHz). The difference of deviation is due to irregularity of cloud and water vapor. As a result, it was found that radiometric distortion in 31 GHz is larger than that in 23 GHz and the value is 1-1.5 K.

(2) S/N evaluation

S/N were evaluated by the following formula:

$$S/N = 10 \log \left(\frac{m}{\sigma} \right) \text{ (dB)}$$

where m is mean and σ is standard deviation of uniform area of ocean in MSR imagery (level 2). Table 2 shows example of S/N evaluation.

Table 2. Example of S/N evaluation

| Frequency (Integration time) | Path 21 | | |
|------------------------------------|----------------|---------------|--------------|
| | April 10, 1987 | Sep. 27, 1987 | Dec. 4, 1987 |
| 23 GHz (10 msec) | 22.0 | 21.1 | 22.5 |
| 23 GHz (47 msec) | 22.0 | 21.2 | 22.6 |
| 31 GHz (10 msec) | 21.3 | 18.4 | 20.5 |
| 31 GHz (47 msec) | 21.4 | 18.5 | 20.5 |

As a result, the following results were obtained:

- 1) The value of S/N in 23GHz is larger than that in 31 GHz by 0.5-2 dB.
- 2) The value of S/N in 47 msec is a bit larger than that in 10 msec. However, the difference is very small.

(3) Dynamic range evaluation

Dynamic range was obtained for MSR imagery (level 2). Example of dynamic range is shown in Table 3.

Table 3. Example of measured dynamic range

| Frequency (Integration time) | July 21, 1987 | | | Dec. 4, 1987 | | |
|------------------------------------|---------------|--------|--------|--------------|--------|--------|
| | DR(K) | MAX(K) | MIN(K) | DR(K) | MAX(K) | MIN(K) |
| 23 GHz (47msec) | 74.6 | 248.3 | 173.7 | 76.0 | 224.9 | 148.9 |
| 31 GHz (47msec) | 82.4 | 239.2 | 156.8 | 77.2 | 221.9 | 144.7 |

$$\text{Dynamic Range (DR)} = \text{MAX} - \text{MIN}$$

These values exist in the dynamic range 300 K of MSR (max: 330 K, min: 30 K). The following results were obtained:

- 1) Dynamic range of MSR imagery in 31 GHz is larger than that in 23 GHz.
- 2) Integration time has no impact on dynamic range.
- 3) Dynamic range in 23 GHz in winter is larger than that in summer. Dynamic range in 31 GHz in summer is larger than that in winter.

2.3 Geometric performance

(1) Geometric distortion evaluation

Position error in MSR imagery (level 2) were measured by using GCPs (island, peninsula etc.) in one path

scene. Measured value of mean of this error is 20 km (23 GHz) and 13 km (31 GHz) which is under the specified value 60 km (23 GHz) and 40 km (31 GHz).

(2) Spatial resolution evaluation

Spatial resolution was measured by using edge of coast of MSR imagery (level 2). Examples of measured value (3dB down) are shown in Table 4.

Table 4. Example of measured spatial resolution

| Frequency (Integration time) | Spatial Resolution (km) (3dB down) | | | |
|------------------------------------|------------------------------------|------|----------------|------|
| | Pixel Direction | | Line Direction | |
| | S.R. | N.N. | S.R. | N.N. |
| 23 GHz (10 msec) | 70.5 | 64.6 | 73.6 | 74.6 |
| 23 GHz (47 msec) | 71.1 | 63.1 | 73.7 | 68.4 |
| 31 GHz (10 msec) | 50.7 | 43.7 | 55.6 | 48.0 |
| 31 GHz (47 msec) | 50.9 | 49.0 | 55.9 | 46.6 |

S.R.: Special Resampling, N.N.: Nearest Neighbors
April 14, 1987 path 25

Spatial resolution depends upon beam width of antenna (1.89 for 23 GHz and 1.31 for 31 GHz) and resampling method to be used. The following results were obtained:

- 1) The antenna pattern correction leads to increase of spatial resolution.
- 2) Resampling N.N. provides higher spatial resolution than special resampling by about 10 %.
- 3) Spatial resolution for 23 GHz in the line direction is worse than that for 31 GHz by 1.4 (ratio of beam widths for 23 GHz and 31 GHz).

3. VERIFICATION OF PHYSICAL MEASUREMENT

3.1 Water Vapor and Liquid Water

One of recurrence equation developed for extraction of water vapor and liquid water under joint research with Communications Research Laboratory (CRL) were installed in MOS-1 data processing system as verification software. Using antenna temperature T_1 (K) for 23GHz and T_2 (K) for 31GHz, water vapor V (kg/m²) and liquid water L (kg/m²) are expressed in the following equation:

$$V = a(T_1 + A_1 - B_1) + b(T_2 + A_2 - B_2) + c$$

$$L = d(T_1 + A_1 - B_1) + e(T_2 + A_2 - B_2) + f$$

$a = a_1(1 + a_2 \cos 2\phi)$ A_1, A_2 : parameter for antenna
 $b = b_1(1 - b_2 \cos 2\phi)$ pattern correction
 $c = c_1(1 + c_2 \cos 2\phi)$ B_1, B_2 : bias error
 $d = d_1(1 + d_2 \cos 2\phi)$
 $e = e_1(1 - e_2 \cos 2\phi)$
 $f = f_1(1 + f_2 \cos 2\phi)$
 ϕ : antenna rotation angle

$$a_1 = 1.0716 \quad b_1 = -0.4862 \quad c_1 = -74.330$$

$$a_2 = 0.0086 \quad b_2 = 0.0180 \quad c_2 = 0.0692$$

$$d_1 = -0.01168 \quad e_1 = 0.03432 \quad f_1 = -3.269$$

$$d_2 = 0.0044 \quad e_2 = 0.0145 \quad f_2 = -0.0481$$

By using this equation, water vapor and liquid water were obtained. Though it is very difficult to measure liquid water as truth data, water vapor can be measured by using radio sonde data. So, comparison between the water vapor obtained by MSR and the water vapor obtained by radio sonde was made as shown in Figure 1. Concerning water vapor extraction estimation, the following results were obtained:

- 1) There was bias error (about 15 kg/m^2) in measured water vapor which is almost constant independent of season. It was found that the bias errors are 21.7K and 18.7K for 23 GHz and 31 GHz , respectively. So, if these bias errors are removed, water vapor is obtained with estimation error about 2.5 kg/m^2 which is a bit smaller than that of Nimbus 7 SMMR data (3.3 kg/m^2).
- 2) When the measured values in Hachijyo-jima (about 200km off the coast of mainland) are compared with those in Chichi-jima (about 1000km off the coast mainland), the influence from radiation of land was found in the data acquired over Hachijyo-jima as shown in Fig.1.

Concerning cloud liquid estimation, the following results were obtained.

- 1) The distribution of cloud liquid obtained from MSR data corresponds with distribution of rain obtained by weather radar. MSR is superior to weather radar on mountain because the former is larger in the area to be observed than the latter.
- 2) Combination of VTIR data (visible and thermal band) and MSR data is useful to observe typhoon and front. Example of front was found in MSR and VTIR data (April 27, 1987, Path 21). Moreover, Typhoon 12 was clearly observed in MSR and VTIR data (August 27, 1987, Path 24).

3.2 Sea ice and snow

NASDA conducted MVP airborne experiment in Feb., 1988 to evaluate observation capability of snow and sea. Airborne MSR data were acquired concerning sea ice and snow on MOS-1 passing days (Feb.9 and 10, 1988).

Concerning sea ice, it was found that there is close relationship between MSR brightness temperature and sea ice concentration obtained by VTIR as shown in Fig.2. Moreover, sea ice distribution can be observed through cloud by MSR. Figure 3 shows the change of drifting ice in the ocean near Hokkaido. MSR data are routinely used in Japan Meteorological Agency to make drifting ice map in ocean near Hokkaido. Capability of all weather observation of sea ice is one of important features for MSR. Combination of MSR, VTIR and MESSR is desirable to observe sea ice. When there is cloud, MSR data can be used. When there is no cloud, wide range observation data (1500km) for sea ice distribution can be obtained by VTIR and high spatial resolution data can be obtained by MESSR.

Concerning snow, MOS-1 MSR data, airborne MSR data and truth data were compared for dry snow in Hokkaido. It was found that there was negative correlation between MSR brightness temperature and snow depth/snow water equivalent as shown in Fig.4. Though there is positive correlation for deep snow, it is necessary to accumulate much data. Moreover, it was found that there is more deviation in 31 GHz than that in 23GHz . This is due to the difference of spatial resolution for 23 GHz and 31 GHz .

4. CONCLUSION

Two years (design life of MOS-1) have passed since the launch of MOS-1 and MOS-1 verification Program (MVP) has been successfully conducted in collaboration with joint research organizations, domestic and foreign MVP participating organizations. As a result of the MVP, MOS-1 observation system was found to show satisfactory condition. In this paper, geometric and radiometric performance of MOS-1 MSR were evaluated from various points of view. Unique properties of MSR were clarified. These results were reflected upon Advanced Microwave Scanning Radiometer (AMSR) which is developed by NASDA and will be mounted on NASA Polar Orbiting Platform (N-POP).

NASDA will continue evaluation until the launch of MOS-1b (Feb., 1990) in collaboration with limited number of cooperative organizations.

ACKNOWLEDGEMENT

The authors would like to thank investigators who participated in the MVP and MSR workshop, particularly Prof. T. Suzuki, University of Electro-Communications, Mr. T. Ojima, Communications Research Laboratory, Dr. A. Shibata, Meteorological Research Institute.

REFERENCES

1. K. Maeda et al., "Geometric and radiometric performance evaluation methods for Marine Observation Satellite-1 Program (MVP)", *Acta Astronautica* vol.15, NO.6/7, pp.297-304, 1987.
2. K. Maeda et al., "Outline of MOS-1 Verification Program (MVP)", 16th ISPRS, Kyoto, July 1-10, 1988, *Geocarto International* (2) pp.13-19, 1988.
3. K. Maeda et al., "Some results of MOS-1 airborne verification experiments-MSR (Microwave Scanning Radiometer)", *Proceedings of IGARSS'87*, pp.1073-1078, Ann Arbor, Michigan, May 18-21, 1987.
4. T. Ojima, K. Maeda and H. Sato, "Evaluations of the MSR antenna pattern and retrieval error of water vapor content", *Proceedings of the third MVP symposium*, pp.89-101, Tokyo, Feb.27-March 1, 1989.
5. A. Shibata et al., "Application of MOS-1 MSR data in Meteorology and sea ice research", *Proceedings of MOS-1 international symposium*, pp.79-89, Tokyo, March 2, 1989.

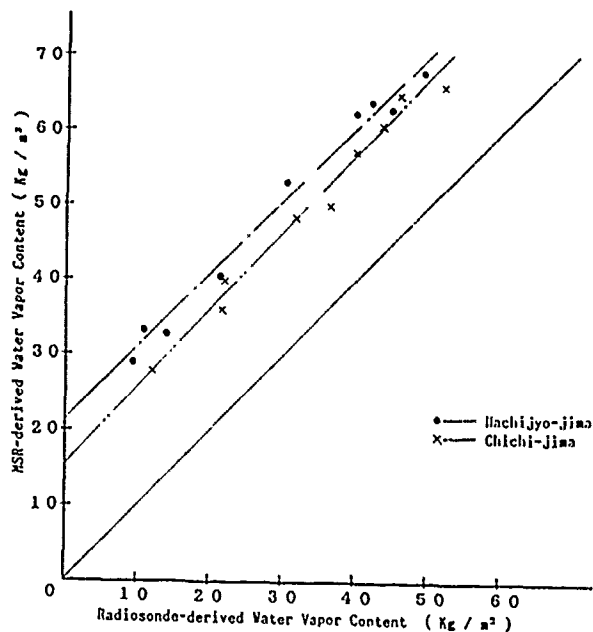


Fig.1 Comparison between water vapor obtained by using MOS-1 MSR data and water vapor obtained by using radio sonde data in the vicinity of two islands: Hachijyo-jima (about 200km off the coast of main land) and Chichi-jima (about 1000km off the coast of main land).

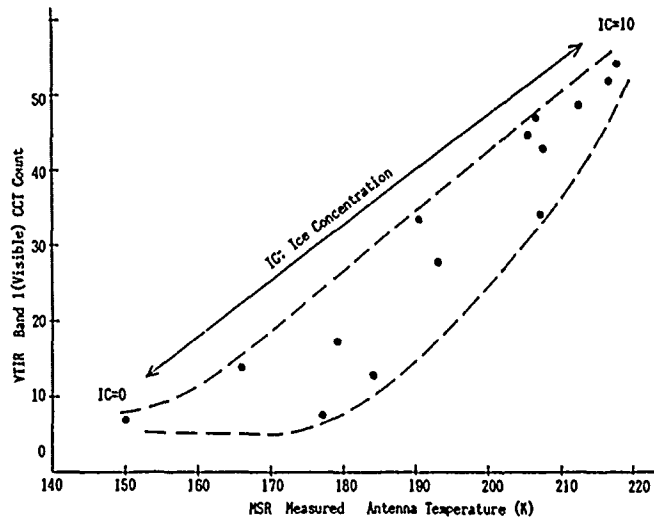
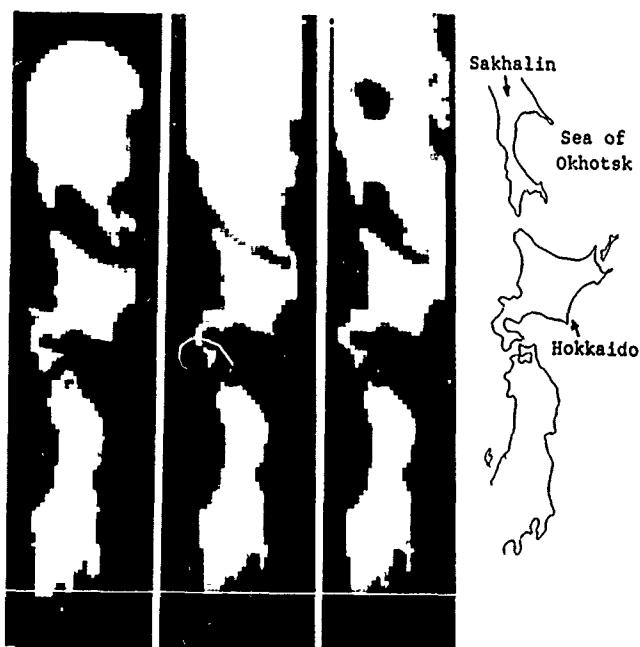


Fig.2 Relationship between MSR antenna temperature and VTIR Band 1 (Visible).



(a)Jan.23,1988 (b)Feb.9,1988 (c)March 31,1988

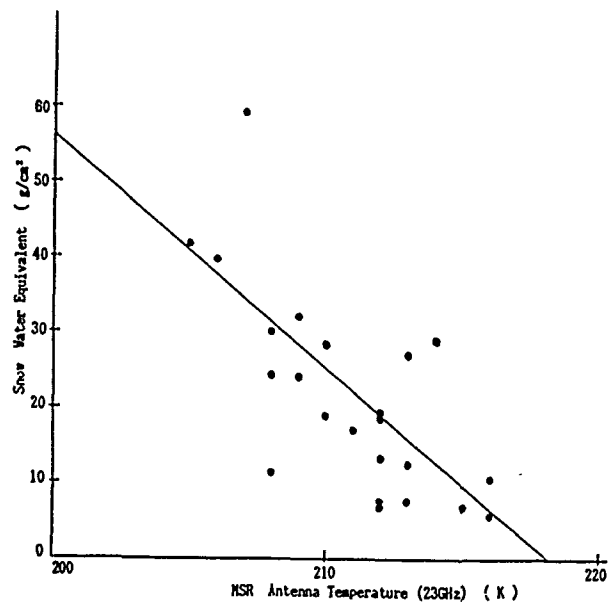


Fig.4 Relationship between MSR antenna temperature and snow water equivalent.

Fig.3 Change of drifting ice in the ocean near Hokkaido. (MSR 31 GHz, path 20, Level 2)

THE CANADIAN MOS-1 PROJECT

T. Butlin, J. Lam, M. Manore and R.A. O'Neill

Canada Centre for Remote Sensing
2464 Sheffield Road
Ottawa, K1A 0Y7 CANADA

ABSTRACT

The Canada Centre for Remote Sensing (CCRS) and the National Space Development Agency of Japan (NASDA) signed an Arrangement under which Canada would receive imagery from 100 passes of the Marine Observation Satellite (MOS-1) over Canada in the period May 1 to October 15, 1988. CCRS assembled a team of researchers from many different groups across Canada to participate in the evaluation of MOS-1 data. The receiving station at Gatineau, Quebec (near Ottawa) was upgraded to track MOS-1 and record the data. Specific target areas and reception dates were suggested by team members and the acquisition was scheduled in collaboration with the NASDA Earth Observation Centre.

A quicklook system was designed and built to catalog the imagery and to write decimated imagery on an analog video disk. The catalog may be searched automatically by a personal computer which then displays the selected Multi-spectrum Electronic Self-Scanned Radiometer (MESSR) images from the video disk. Raw data from all three sensors carried by MOS-1 may be transcribed by the quicklook system from the High Density Digital Tape recorded at the ground station to Computer Compatible Tapes (CCTs) for input to geometric and radiometric correction software.

1. INTRODUCTION

The National Space Development Agency of Japan (NASDA) launched the Marine Observation Satellite (MOS-1) on February 19, 1987. The satellite carries three imaging sensors of interest to the remote sensing community: the Multi-spectrum Electronic Self-Scanning Radiometer (MESSR) is a push-broom imager operating in four bands in the visible and near infrared portion of the spectrum; the Visible and Thermal Imaging Radiometer (VTIR), a scanner of lower spatial resolution imaging in the visible and thermal infrared portion of the spectrum; and the Microwave Scanning Radiometer (MSR), a two frequency imaging passive microwave radiometer. This sensor complement allows the surface of the earth to be imaged simultaneously in three separate spectral regions. In the past, it was necessary to use images from two or three different satellites making it difficult to understand relationships between the various spectral regions. The simultaneity of imagery is of particular importance in the study of ocean features which frequently change significantly in a matter of a few hours.

2. THE ARRANGEMENT

Following the exchange of enabling letters between the Ministers responsible for the MOS-1 Programs in Canada and Japan, the Canada Centre for Remote Sensing (CCRS) and NASDA signed an Arrangement for the Reception of Data from the Marine Observation Satellite. CCRS was permitted to select and track the MOS-1 satellite for 100 orbits over Canada in the period May 1 to October 15, 1988. Because of the reception circles of the Canadian ground stations cover most of North America, imagery of both Canada and the U.S.A. could have been received under the Arrangement.



Figure 1. Black and white rendition of a MESSR image centred on Ottawa. The Gatineau Satellite Station is slightly below the centre of the image on the Gatineau River (the small river flowing from northwest to southeast).

The Arrangement also permitted the Canadian MOS-1 Project Team (see section 3) to use the MOS-1 imagery: to evaluate the data itself (sensor performance, technical characteristics, etc.), to develop processing algorithms and to demonstrate applications. Because the MOS-1 Program was seen, at the time, to be a development program by both CCRS and NASDA, the Arrangement did not permit CCRS to distribute the data. Data for purposes outside those of the research interests of the Canadian MOS-1 Project Team was to be obtained through NASDA (see section 9).

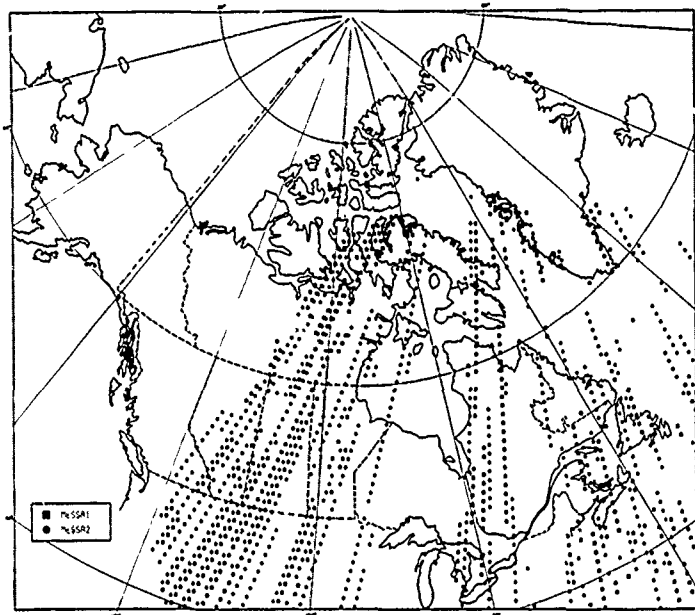


Figure 2. A map of Canada with provincial boundaries showing the centres of 'cloud-free' MESSR scenes acquired in the period May 4 to December 1, 1988. MESSR1 scenes are indicated with squares, MESSR2 with circles.

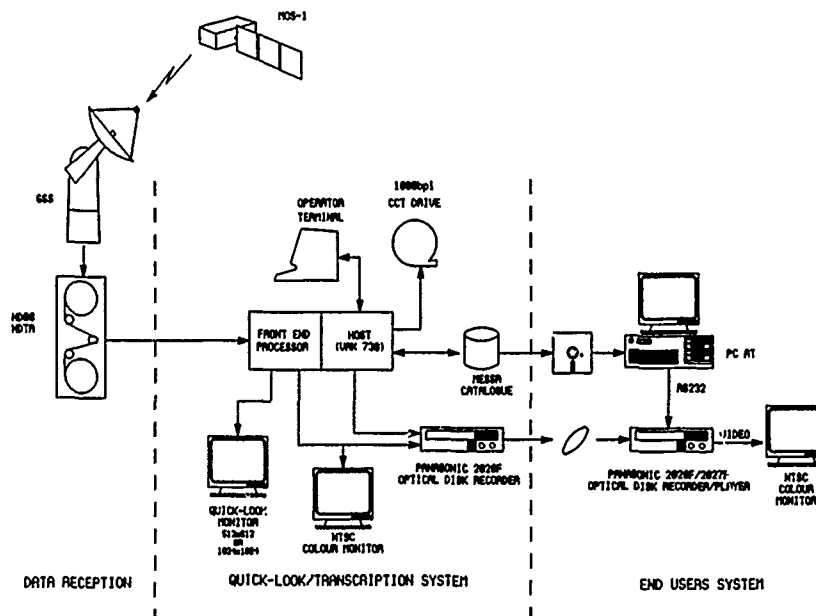


Figure 3. Flow of MOS-1 data from reception and recording, through quicklook and catalog generation, to the search of the inventory.

3. THE CANADIAN MOS-1 PROJECT TEAM

CCRS invited almost all active Canadian research groups to participate in the Canadian MOS-1 Project Team. Most of the groups proposed applications development projects exploiting imagery from the sensors on the satellite. A few proposals were received to acquire sample data sets that could be used to develop processing or image analysis software. As each of the proposals addressed a different aspect of the MOS-1

data, no proposals were rejected.

There were 30 team members. CCRS undertook to acquire the data required by each of the team members. The requested data were acquired successfully in almost every case. Raw data have been delivered in both photographic and digital form to some team members; most team members needed processed imagery, deliveries of which did not start until late April 1989 (see Section 7).

MOS-1 team members were informed of progress and status of the overall project by means of an informal newsletter. In addition, a team meeting was held in Ottawa in early December 1988 to allow members to examine the catalog of quicklook imagery.

4. THE RECEPTION OF MOS-1 DATA

The Gatineau Satellite Station (GSS) at Cantley, Quebec (near Ottawa) was upgraded to track MOS-1. GSS began to receive MESSR and VTIR imagery on May 4, 1988 and all three sensors were recorded beginning June 1, 1988. A total of 169 orbits were tracked from May 1 to December 31, 1988. Due to a problem with the satellite, there are four paths over Canada for which the satellite cannot be turned on. As it has turned out, the areas accessible from these paths have not contained targets of particular interest to the Canadian MOS-1 Project Team. A map of Canada locating the 'cloud-free' scenes which have been acquired is shown in Figure 2. About 1400 scenes are identified on this map of the 2116 'cloud free' scenes of Canada and the United States acquired by the Gatineau Satellite Station.

5. MOS-1 QUICKLOOK AND TRANSCRIPTION SYSTEM

A system was required to monitor the condition of the satellite downlink in real time, to generate quicklook imagery and to catalog data for the MOS-1 sensors. Earlier such systems had written monochrome quicklook imagery to film or microfiche; the MOS-1 system was to write colour quicklook images on an analog video disk. Cloud cover estimates were to be entered in the catalog by an operator working off-line, using a video display of the MESSR imagery. A second major function was to transcribe raw scenes of MESSR, VTIR and MSR, as well as telemetry data from High Density Digital Tape to Computer Compatible Tapes (CCTs). There was also a need to integrate the MESSR catalog with the SPOT and LANDSAT catalogs, and to provide the capability of searching the catalog on a personal computer (PC).

5.1 Implementation

An existing SPOT Quicklook System was modified to process MOS-1 data. The host computer is a VAX-11/730, with the usual peripherals and some custom hardware to support the MOS-1 requirements. It has 4 Mbytes of memory, a floating point accelerator, a total of 572 Mbytes of disk space, a 1600 bpi tape drive and an 8-line communication interface that connects terminals, a data link and an optical disk recorder. A block diagram of the data flow for MOS-1 is shown in Figure 3.

MOS-1 data, after being received and recorded on a Honeywell HD96 High Density Tape Recorder (HDTR), are input to the VAX computer via the custom Front-end Processor (FP). The FP synchronizes, sub-samples and reformats the satellite imagery data, and controls the colour display. Two display controllers are available, a Revolution Series (Number 9 Corporation) graphic display board produces NTS-C video imagery while a Metheus Omega 530 display controller generates a 512x512 or 1024x1024 pixel video display. No radiometric or geometric corrections are applied to the data before display.

A Panasonic TQ-2026F video optical disk recorder (ODR) records each raw and un-annotated MESSR image in NTS-C format directly from the FP. The address of the ODR frame to which the image is written is controlled by the host computer. Up to 24000 scenes or about 350 Canadian passes can be recorded on a single-sided 20 cm

diameter optical disk.

The performance of the MOS-1 Quicklook and Transcription System is summarized in Table 1.

Table 1. Summary of the MOS-1 Quicklook and Transcription System Throughput.

| | |
|-----------------------|--|
| MESSR display speed | twice real-time |
| MESSR quicklook speed | twice real-time or less than an hour when batched to process an HDDT of up to 6 passes |
| Cloud cover entry | less than 5 minutes per pass |
| MESSR scene-to-disk | about 2 minutes |
| VTIR scene-to-disk | about 3 minutes |
| MSR scene-to-disk | about 5 minutes per pass |
| Disk-to-CCT | about 40 minutes on VAX-11/730 or about 10 minutes |

5.2 Quicklook and Catalog Generation

For catalog generation, MESSR data are read from the HDT, decimated by the FP and displayed in NTS-C video format. The control process in the host computer uses the pre-generated nominal World Reference System (WRS) framing information and the station time code (from a separate track on the HDT) to determine the scene end time which triggers the recording of the displayed scene to the ODR. At each scene end time, catalog information including WRS indices, station time, scene quality information and ODR frame number is written to the catalog file. Catalog files are generated in the same free-format as is used for SPOT thereby allowing easy integration of the MESSR files into the satellite inventory.

After the quicklook recording process, an operator assesses the recorded pass and estimates cloud content of each scene. Two catalogs are generated for each pass, the 'master' containing information of all scenes, and the 'cloud-free' containing information only of scenes with 20% or less cloud cover.

Using the master file of MESSR imagery on the ODR and the catalog, a subset of the imagery may be selected and written to VHS video cassettes. All team members have been provided with a video tape of the 'cloud-free' catalog.

5.3 CCT Production

During CCT production, the data flow from the HDTR to the CCT drive is controlled by a single supervisory program in the host computer. On input, the HDT is played from slightly before the start of the required scene, the bit stream is synchronized and formatted by the FP and the selected sensor data is sent to intermediate storage on disk. At the same time as the data are being sent to disk, the data are sub-sampled by the FP and the decimated image is sent to the Metheus display. Following this input process, the integrity and quality of the data on disk are analyzed by the computer in preparation for output to CCT. The output CCT format follows the NASDA specification.

When MSR and telemetry data are being extracted, a whole pass may be processed at once; for the other sensors, one scene is processed at a time.

5.4 Inventory Search on a Personal Computer

A simple and convenient means for accessing MESSR quicklook image inventory has been developed as part of the MOS-1 project. It was designed to demonstrate a video optical disk used interactively with the image catalog as a replacement for microfiche, both at the order desks and at the sites of sophisticated end-users.

All MESSR scenes received at the Gatineau Satellite Station were recorded on optical disk (Section 5.2). The catalog information collected by the MOS-1 Quicklook System and stored in the satellite inventory was downloaded to the hard disk of a PC which in turn controlled the ODR through an RS-232 interface. Due to the storage restrictions on PCs and unsophisticated search mechanisms, only the 'cloud-free' versions of the catalog files stored on the VAX computer are normally transferred to the PC.

Simple commands were implemented enabling the user to select a data set for input, generate a subset, select and display individual scenes as well as store and retrieve subsets. Standard data sets available are MOS-1 or SPOT, master or 'cloud free'. The selected input data set file may be further reduced to a subset according to a range of location (path/row or longitude/latitude) or time period (year/day) specified by the user. Following the generation of a subset, its size and the first available scene are automatically displayed. Each scene in the subset may be selected from the optical disk for display on a separate NTS-C video monitor while the corresponding catalog information are shown on the PC screen. A user may use single-key cursor commands to 'walk-through' the entire subset of up to 750 scenes.



Figure 4. Black and white rendition of a MESSR scene of the western end of Lake Ontario. Toronto is in the top right quadrant of the image, the city of Hamilton is at the centre of the lower edge.

The catalog system represents a major technological advance and a potential improvement in

the service which will be offered to all Canadian users of remotely sensed imagery when a similar capability is adopted for the LANDSAT and SPOT image catalogs.

6. PRELIMINARY EVALUATION OF MOS-1 IMAGERY

(Manore and Butlin, 1989) have carried out a preliminary analysis of the raw MOS-1 imagery to determine which corrections are must be applied in the processing of the data prior to its delivery to team members wishing to carry out applications research. Manore and Butlin found that, in addition to the inter-detector calibration, the MESSR images require significant geometric correction to register the four spectral channels. The raw VTIR data was found to be most satisfactory and in need of little additional processing. The radiometric artifacts apparent in the raw MSR imagery will need to be corrected using the sensor calibration information.

7. PROCESSING

MacDonald, Dettwiler and Associates (MDA) (Erickson, Robertson, Sharpe, 1989) of Vancouver is developing, on behalf of the Canadian MOS-1 Project Team, a software package to process MOS-1 imagery. The software will accept raw CCTs from the MOS-1 Quicklook and Transcription System and apply the appropriate radiometric and geometric corrections. As this will be a prototype system, the through-put will be limited to only a few scenes per day. MDA will, in the course of the development, process 20 scenes from each of the MOS-1 sensors. The first MESSR scenes are expected in late April 1989. In June 1989, the software will be delivered to CCRS so that additional scenes can be processed for the team members.

8. APPLICATIONS BEING INVESTIGATED

Several typical full scenes of MESSR imagery (105 km square) are shown in Figures 1, 4 and 5. Raw CCTs of these scenes were produced on the MOS-1 Quicklook and Transcription System and then imaged on a FIRE 240 film recorder.



Figure 5. A black and white rendition of a MESSR scene of Anticosti Island in the Gulf of St. Lawrence. Ocean features are apparent near the shore. Variations in forest cover and bogs are clearly visible.

The applications of MOS-1 imagery being investigated by the Canadian MOS-1 Project Team include physical oceanography, sea ice studies, forest clear cut mapping and inventory, forest fire mapping, agriculture and land use, topographic map revision, geology and hydrocarbon exploration. These are the principle disciplines in which other forms of satellite imagery are also used in Canada. Many of the studies focus on evaluating MOS-1 as an alternative source of data, particularly as the frequency of coverage provided by other satellites alone is inadequate for many studies, especially when cloud cover is been taken into account.

Most of the studies carried out by the Canadian MOS-1 Project team are integrated into existing programs. In many cases imagery was collected of well established test sites supported by various federal, provincial and university laboratories. In the case of one oceanographic experiment, MOS-1 imagery was acquired coincident with a major survey around Georges Bank, off the coast of Nova Scotia.

The following is one example of the type of project in which MOS-1 data is being used. Only a small group of scientists at the Atmospheric Environment Service (AES) have had experience with passive microwave imagery. Thus, to introduce this important data type into a wider group of applications, CCRS and AES will carry out a collaborative project with Ph.D. Associates of Toronto, one of the few companies in Canada having analyzed this type of data in the past. Existing MSR scenes of the ocean will be used to develop a 'calibration' of the MSR through a comparison with the SSM/I sensor carried by the Defense Meteorology Satellite. Additional MSR imagery is being collected this winter over the Saskatchewan prairies and southern Ontario where snow thickness surveys are carried out as part of a continuing hydrometeorology project. Standard algorithms will be used to reduce the MSR brightness temperatures and map the water equivalence of the snow cover. These results will be compared to the surface measurements.

9. CONCLUDING REMARKS

An amendment to the Arrangement has been negotiated whereby CCRS may continue to receive MOS-1 data. In addition, CCRS is permitted to distribute MOS-1 data from its archives.

CCRS will upgrade the facilities at Prince Albert Satellite Station to receive MOS-1, to give nearly complete North American coverage, beginning in the Spring of 1989. This is seen as an important step in meeting the demands of the user community and encouraging the wider use of MOS-1 data.

This paper was also presented at the MOS-1 International Symposium (Butlin *et al.*, 1988).

10. REFERENCES

1. Butlin, T., J. Lam, M. Manore and R.A. O'Neil, "The Canadian MOS-1 Project", presented at the MOS-1 International Symposium - Utilization of MOS-1 Data and Earth Observation in the Future, March 2, 1989, Tokyo, Japan.
2. Manore, M. and T. Butlin, "Preliminary Evaluation of Typical Canadian Scenes of MOS-1 Imagery", Proceedings of 3rd Symposium on the MOS-1 Verification Program, February 28 to March 1, 1989, Tokyo, Japan.

3. Erickson, A., B. Robertson, and B. Sharpe, "Geometric and Radiometric Correction of MOS-1 Imagery in a Canadian Processing System", IGARS '89 Symposium, 1989.

The three MOS-1 MESSR scenes reproduced in this paper are copyright NASDA 1988 and produced by CCRS under a NASDA license.

4. Figure 1. Sensor: MESSR
Acquired: November 15, 1988
Scene centre: 45.51°N, 75.11°W
WRS: 164 - 056
Scene ID: 10623 - 154920
5. Figure 2. Sensor: MESSR
Acquired: May 31, 1988
Scene centre: 43.37°N, 78.99°W
WRS: 166 - 59
Scene ID: 10455 - 155552
6. Figure 3. Sensor: MESSR
Acquired: June 24, 1988
Scene centre: 49.19°N, 61.37°W
WRS: 156 - 51
Scene ID: 10479 - 145420

GEOMETRIC AND RADIOMETRIC CORRECTION OF MOS-1 IMAGERY IN A CANADIAN PROCESSING SYSTEM

Arvon Erickson

Canada Centre for Remote Sensing
2464 Sheffield Road, Ottawa, Ont.
Canada K1A 0Y7

Brian Robertson and Kelly Wiebe

MacDonald Dettwiler & Associates Ltd.
13800 Commerce Parkway, Richmond, B.C.
Canada V6V 2J3

ABSTRACT

The Marine Observation Satellite (MOS-1) has been operational and routinely acquiring data for over 20 months. As part of the MOS-1 verification project, the Canadian Centre for Remote Sensing (CCRS) has acquired over 160 orbits of MOS-1 data. In order to effectively evaluate this data, it must first be processed to correct for radiometric and geometric errors. To meet this need, MacDonald Dettwiler has developed a prototype MOS-1 correction system capable of radiometrically and geometrically correcting raw MOS-1 data. This paper reviews the Canadian MOS-1 processing capabilities. It begins with a review of the MOS-1 sensors and image characteristics. Following this, the radiometric and geometric correction algorithms of the Canadian prototype system are described. Finally, examples of the processed imagery are presented.

Keywords: MOS-1, Geometric Correction, Radiometric Correction.

puter compatible tapes (CCTs) generated by the CCRS Transcription System. The raw data is absolutely radiometrically calibrated, geometrically corrected to a map projection, and output on LOGSWG format CCTs. Geometric processing resamples the image to a map projection correcting for sensor alignment, band-to-band registration, earth curvature and rotation effects, satellite orbit and attitude, and temperature dependent sensor characteristics. Radiometric processing converts the raw digital numbers (DNs) to physical units correcting for temperature dependent sensor characteristics using both inflight sensor calibration data and preflight information.

This paper describes the Canadian processing of MOS-1 data. The three MOS-1 sensors are reviewed with emphasis on the quality of the raw image data. The radiometric and geometric correction algorithms of the Canadian prototype system are then discussed followed by which examples of processed imagery are presented.

1 INTRODUCTION

Since its launch on February 19, 1987, the Marine Observation satellite (MOS-1) has been routinely acquiring data over North America. As part of the MOS-1 verification project, the Canadian Centre for Remote Sensing (CCRS) has tracked over 160 orbits of MOS-1 data from which test scenes will be chosen for sensor evaluation and application studies. Preliminary evaluation of the raw data imagery generated by the CCRS Quicklook Transcription System [Butlin 89] indicated that in order to effectively analyze the data, it would first have to be processed to correct for radiometric and geometric errors. The goal of the current project is to provide corrected data for Canadian researchers to assess the quality and utility of MOS-1, and provide CCRS with limited MOS-1 processing capabilities to meet further processing needs on an interim basis.

To meet this goal, MacDonald Dettwiler has developed a prototype radiometric and geometric correction system capable of processing data from all three MOS-1 sensors. The system, which is based on MacDonald Dettwiler's Geocoded Image Correction System (GICS) engine, takes as input raw MOS-1 data on LOGSWG format com-

2 IMAGE CHARACTERISTICS

MOS-1 was designed to observe and monitor land, ocean, and atmospheric processes. The satellite carries three sensors which observe the earth at a variety of resolutions in the visible, near-infrared, thermal infrared, and microwave regions of the electromagnetic spectrum. In this sense, MOS-1 provides the remote sensing community with an ideal opportunity to study and compare data from a diverse set of sensors. As data is simultaneously acquired by all three sensors, the traditional problems associated with comparing multisensor data taken at different times and under different environmental conditions are eliminated.

2.1 MESSR IMAGERY

The Multispectral Electronic Self-Scanning Radiometer (MESSR) was designed to observe the earth using four spectral bands in the visible and near infrared spectrum. There are actually two MESSR sensors, systems 1 and 2, which point at fixed angles (nom-

| MESSR System | Misregistration From Band 1 | | |
|--------------|-----------------------------|------------|------------|
| | Band 2 | Band 3 | Band 4 |
| 1 | 1.07/0.72 | 2.25/-0.28 | 0.80/-0.86 |
| 2 | -0.12/0.58 | 4.87/1.38 | 1.05/1.32 |

Table 1: Band-to-band misregistration in raw MESSR data. The values are given in the along and across track directions and are in units of pixels.

inally 2.73°) to either side of nadir. Under normal operating conditions, imagery may be acquired from only a single system at a given time. Each MESSR is a push broom sensor incorporating a single 2048 element CCD linear array in each spectral band. It has a nominal ground resolution of 50 meters acquiring data over a 100 kilometer swath to one side of the satellite track. Radiometrically, the MESSR's four spectral bands are similar to those of the Landsat MSS sensor. This, together with the similar geometric resolutions, has sparked interest in comparing the quality of data acquired by the two sensors.

Evaluation of raw MESSR imagery has shown that the data contains a number of radiometric and geometric artifacts which hamper its analysis [Manore 89, Henry 89a]. Significant band-to-band misregistration, caused by misalignment of the four detector arrays, is present in images acquired by both systems. The magnitude of this misregistration varies between system 1 and system 2, and may be as large as 6 pixels. In many applications such as land use classification, accurate band-to-band registration is critical and any misalignment can introduce significant misclassification errors. An analysis of six MESSR images was undertaken to characterize the magnitudes of the detector array displacements and their dependence on temperature. Using Band 1 as a reference, the spectral bands were correlated against one another at uniformly distributed grid points over each image. Table 1 tabulates the magnitudes of the misregistration found. These results are consistent with those reported in the literature [Manore 89, Henry 89a].

Analysis of the radiometric quality of raw MESSR data has identified a single major artifact. Vertical striping, typical of linear array sensors, has been identified in all MESSR images studied thus far. This striping takes two forms: random striping caused by differences in the gains and dark current offsets of adjacent CCD elements, and periodic striping caused by the even-odd shift registers of the CCD. The random striping may be removed to a large extent by absolutely calibrating the data using NASDA supplied calibration coefficients. The periodic even-odd striping is not as easily removed.

2.2 VTIR IMAGERY

The MOS-1 Visible and Thermal Infrared Radiometer (VTIR) is designed to observe the temperature of earth surfaces (sea surfaces) and cloud tops, ice distribution and cloud distribution using one band in the visible region and three bands in the infrared region. The VTIR is similar to the NOAA AVHRR sensor, although it does have a number of distinct characteristics. Geometrically, the VTIR is a mechanical scanning radiometer consisting of a single detector in each band. The ground resolution of the VTIR is nominally 900 meters in the visible band, and 2.7 kilometers in the infrared bands. Radiometrically, the VTIR provides 16 gain modes in each band. Calibration of the data is accomplished by observing a built-in black body and deep space when the scanning mirror is not sweeping the earth.

The image quality of raw VTIR data is relatively good with only a few problems being reported in the literature. While the band-to-band registration is much better than for MESSR, there is still a 1

to 2 pixel misalignment between the different detectors. Radiometric artifacts may also be observed in the data in the form of horizontal vertical striping noise almost periodic in nature.

2.3 MSR IMAGERY

The MOS-1 Microwave Scanning Radiometer (MSR) is designed to observe snow conditions, the amount of vapor in the atmosphere over oceans, and the amount of water in cloud and sea ice by receiving microwave noise signals emitted through the air from ground surfaces and oceans. The MSR employs a conical scan system which employs an offset cassegrain antenna common to both bands. Half the time required for one rotation of the antenna is used for earth observation while the remaining half is used for recording the calibration data of low and high reference temperature sources. The MSR produces data in two bands (23 GHz and 31 GHz) at two different integration times (47 msec and 10 msec) at a radiometric resolution of 10 bits. The ground resolution is nominally 32 kilometers in the two 23 GHz bands and 23 kilometers in the two 31 GHz bands.

Because of the conical scan system, the ground scanning pattern of the MSR is semicircular, leading to large geometric distortions in the raw imagery. Analysis of the radiometric quality of this data has shown the presence of horizontal striping in the imagery as well as random pixel dropouts, particularly prevalent near the top and bottom portions of the imagery. As the processing unit of MSR data is a full satellite pass, these dropouts are most likely caused by errors in communication with the satellite when it is near the horizon.

3 RADIOMETRIC & GEOMETRIC CORRECTION

To allow for overlay and comparison of images from the different MOS-1 sensors, the raw data must first be corrected for geometric distortions introduced by the nature of the imaging geometry. Further, the imagery must be absolutely radiometrically calibrated through the conversions of raw DN's to physically relevant quantities so as to facilitate meaningful comparisons of multitemporal and multisensor data. This is necessary to allow the effective utilization and comparison of datasets from all three MOS-1 sensors both amongst themselves and with other sensors such as the Landsat MSS and TM sensors, and the NOAA AVHRR sensor.

To meet these needs, MacDonald Dettwiler has developed, on behalf of the Canadian MOS-1 Project Team, a prototype MOS-1 processing system capable of geometrically correcting data from all three MOS-1 sensors to a standard map projection, and absolutely calibrating the raw DN's to physical units. The system is based on MacDonald Dettwiler's GICS engine with enhancements to handle the MOS-1 specific satellite and sensor systems. The corrected imagery is output on LOGSWG compatible CCTs.

Geometric correction is performed in a two stage process. In the first stage, the correspondence between any given pixel in the input imagery and a point on the earth's surface is established. This correspondence is called the *forward transformation*. In the second stage, the input imagery is resampled to a regular grid using a user selectable resampling kernel. The processing corrects for sensor alignment, band-to-band registration, earth curvature and rotation effects, satellite orbit and attitude, and temperature dependent sensor characteristics. The primary output of the system is an absolutely radiometrically corrected and systematically georeferenced product. Systematic refers to the fact that the images are geometrically corrected using only *a priori* information such as the sensor geometry, ephemeris and attitude data; no ground truth is required. Georefer-



Figure 1: A geometrically and radiometrically corrected MESSR Band 3 image of the Bay of Fundy acquired on September 6 1988. The data was acquired using high gain mode by MESSR System 2. For comparison, a small subimage taken from the processed VTIR image shown in Figure 2 is given in the inset. (Copyright NASDA 1989)

enced products are in a map projection oriented in the direction of the nominal satellite heading at the scene center. Output pixel spacings for the three sensors have been chosen such that they differ by integral multiples from other sensors processing by the GICS system allowing for easy multisensor image overlay and comparison.

Radiometric correction is also performed in a two stage process. Input pixels are first converted to physical units (typically radiance) using telemetry data, as well as inflight and preflight calibration data. The corrected values are then converted back to digital numbers for output to CCT using either fixed or variable gains and offsets. These parameters are stored in the radiometric ancillary record of the corrected CCT, thus may be used to convert the corrected DN's back to physical units.

3.1 MESSR PRODUCTS

The standard processing of MESSR data generates products with 50 meter pixel spacings which have been systematically georeferenced to either the Universal Transverse Mercator (UTM) or Lambert Conformal Conic (LCC) map projections. The process corrects for sensor alignment, band-to-band registration, earth curvature and rotation effects, satellite orbit and attitude and temperature dependent distortions of the CCD arrays.

An analysis of the geometric accuracy of the systematically corrected MESSR products was performed on a MESSR System 1 image of Hamilton, Canada acquired on September 28 1988 using seventeen control points distributed uniformly throughout the image. After removal of the along and across track bias errors (due to the systematic nature of the spacecraft modelling), the combined RMS error in locating accuracy was found to be 44 meters, i.e. less than one pixel error.

The scale accuracy of the product was measured using the same 17 GCPs. Scale accuracy is defined as the error between actual distances on the ground, d_g , and the distances estimated from the imagery, d_i

$$\text{scale accuracy} = \frac{d_g - d_i}{d_g} \quad (1)$$

The scale accuracy of the corrected MESSR data was found to have an RMS value of 0.001, which is similar to the values reported for SPOT data by CNES [Henry 89a].

The radiometric correction of MESSR data includes absolute calibration of the raw DN's to radiance using satellite telemetry data and preflight measurements supplied by NASDA, relative calibration to remove residual striping, and decompression of the 6-bit data. The calibrated radiance is converted to an 8-bit DN using fixed band/gain mode dependent radiance limits before output to CCT. As MOS-1 has no in-flight absolute calibration capability, one can not directly monitor and track the radiometric degradation of the sensor over time. However, work by Henry *et. al.* [Henry 89b] comparing MESSR imagery with near simultaneous SPOT MLA data has shown that it is possible to achieve absolute radiometric calibration with an uncertainty as low as 10 percent.

An example of a geometrically and radiometrically corrected MESSR System 2 image of the Bay of Fundy is shown in Figure 1. The image was acquired using high gain mode on September 6 1988. To demonstrate the utility of geometrically the data, a small subimage of a corrected VTIR image over the same area is shown. Comparison of the processed data is made easier since both images have been corrected to the same map projection, and the pixel sizes are integral multiples of one another.

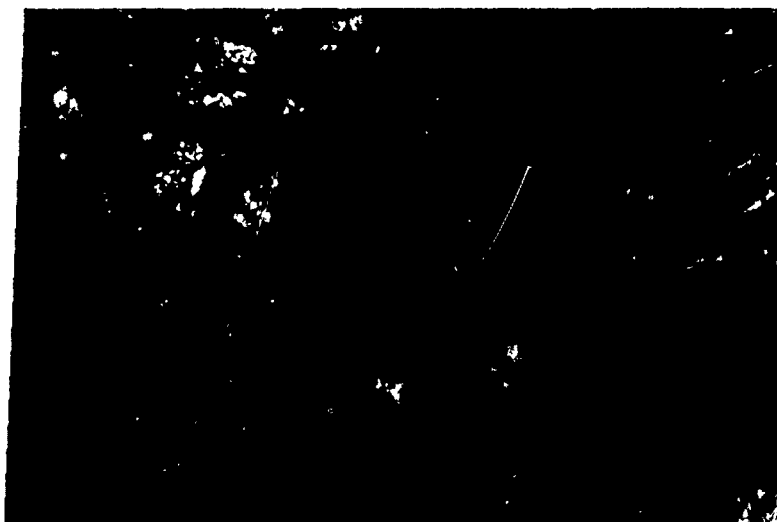


Figure 2: Part of a geometrically and radiometrically corrected VTIR visible band (Band 1) image off the east coast of Canada acquired on September 6 1988. The image has been corrected to the Lambert Conformal Conic map projection with 550 meter pixel spacings. (Copyright NASDA 1989)

3.2 VTIR PRODUCTS

Standard processing of VTIR data generates systematically georeferenced products in the LCC map projection which have been corrected for band-to-band misregistration, earth curvature and rotation effects, panoramic distortions, satellite orbit and attitude errors, and different spatial resolutions between bands. The corrected data is re-sampled to a 550 meter pixel spacing to maintain the pixel sampling rate of the visible band while at the same time allowing easy comparison with 1100 meter AVHRR data. The raw DN's are converted to radiance using both inflight calibration data obtained by viewing a reference black body and deep space, as well as preflight data. The corrected radiance is converted to an 8-bit DN using scene specific gains and offsets before being output to CCT.

An example of a geometrically and radiometrically corrected VTIR product off the east coast of North America acquired on September 6 1988 is shown in Figure 2. Geometric processing has corrected for the uneven line and pixel spacing, differing resolution, and the large panoramic distortions present in the raw data. This allows for easy comparison with other sensors as was demonstrated in Figure 1.

3.3 MSR PRODUCTS

Because of the unique characteristics of the MSR sensor, two types of products may be produced by the MOS-1 processing system. To meet the requirements of specialized users, the system is capable of generating an ASCII table from the raw MSR data which tabulates the location (in LCC and geodetic coordinates), the temperature, and the solar conditions for each observation point. To meet the needs of more traditional image processing users, the system is also capable of generating raster MSR products. In both cases, the system corrects geometric errors due to earth curvature and rotation effects, satellite orbit and attitude errors, different spectral resolutions, and the sensor scanning characteristics. Radiometric correction of MSR data converts the raw DN's to brightness temperature using inflight calibration information. The correction has been further supplemented to allow the detection and correction of the random pixel dropouts.

4 CONCLUSIONS

Canadian MOS-1 data processing capabilities have been described. A prototype system has been developed by MacDonald Dettwiler for CCRS capable of radiometrically and geometrically correcting data from all three MOS-1 sensors. The processing capabilities include:

- systematic geometric correction to a map projection modelling and removing distortions introduced by the sensor, satellite, and earth;
- absolute and relative radiometric corrections;

The system will provide CCRS with the capability to meet the processing requirements of Canadian MOS-1 Project Team, and to meet additional MOS-1 processing needs on an interim basis.

References

- [Butlin 89] T. Butlin, J. Lam, M. Manore, and R. A. O'Neil. The Canadian MOS-1 project. In *Marine Observation Satellite-1 International Symposium*, pages 180-189, National Space Development Agency of Japan (NASDA), 1989.
- [Henry 89a] P. Henry and G. Begni. MESSR image quality. In *The Third Symposium on MOS-1 Verification Program*, page 435, Earth Observation Center, National Space Development Agency of Japan, 1989.
- [Henry 89b] P. Henry and G. Begni. MOS-1 absolute calibration using SPOT1 as an intermediate radiometer. In *The Third Symposium on MOS-1 Verification Program*, page 436, Earth Observation Center, National Space Development Agency of Japan, 1989.
- [Manore 89] M. Manore and T. Butlin. Preliminary evaluation of typical Canadian scenes of MOS-1 imagery. In *The Third Symposium on MOS-1 Verification Program*, pages 191-200, Earth Observation Center, National Space Development Agency of Japan, 1989.

A STUDY ON SEA ICE MONITORING USING MOS-1/MSR

K. Cho & K. Takeda

Remote Sensing Technology Center of Japan(RESTEC)
7-15-17, Uni-Roppongi Bldg., Roppongi, Minato-ku, Tokyo, Japan

K. Maeda & H. Wakabayashi

Earth Observation Center(EOC)
National Space Development Agency of Japan (NASDA)
Hatoyama, Saitama-ken, Japan

ABSTRACT

In order to verify the capability of the MOS-1 Microwave Scanning Radiometer(MSR) for sea ice monitoring, airborne experiments were carried out by NASDA in 1985 and in 1988 in the Okhotsk Sea. Both airborne and MOS-1 MSR data were analyzed in this study. Ice thickness and ice concentration were estimated from the simultaneously corrected Visible and Thermal Infrared Radiometer(VTIR) data and were compared with the MSR data. The airborne MSR brightness temperature has shown a good linear relationship with sea ice thickness. More than 30K brightness temperature difference was observed between open water and ice floes. The tendency of the increase of ice concentration with the increase of MSR brightness temperature was observed. However, the airborne MSR brightness temperature of thin ice floes were mostly overlapped with ice-water mixed area. The MOS-1 MSR brightness temperature has shown a linear relationship with the sea ice concentration estimated from the MOS-1 VTIR.

Keywords: Passive Microwave Sensor, MOS-1, MSR, VTIR, Sea Ice Concentration

1. INTRODUCTION

Satellite passive microwave sensors provide us useful sea ice information with wide coverage, through cloud, and both day and night. However, as the footprint size of these sensors are usually very large (more than 20km), one footprint is expected to be a mixture of a various types of sea ice and open water. Therefore it is not easy to evaluate the relationship between sea ice condition and satellite passive microwave sensor data.

One of the advantage of MOS-1 is that, it can observe same phenomena at the same time with the three different sensors which are the

Multi-spectrum Electronic Self-scanning Radiometer(MESSR), the Visible and Thermal Infrared Radiometer(VTIR) and the Microwave Scanning Radiometer(MSR). The 31GHz channel of MSR has good atmospheric penetration, and expected to provide useful information for sea ice monitoring. While VTIR visible channel provide us more detailed image of the sea ice under the fine weather. Assuming that the detailed sea ice condition can be estimated by the VTIR data, one can easily evaluate the relationship between the MSR data and sea ice condition by comparing the MSR data with the VTIR data. From this point of view, the airborne data and the satellite data were analyzed to evaluate the capability of MSR for sea ice monitoring.

2. Airborne Data Analysis

As part of the MOS-1 Verification Program, a series of airborne experiment for sea ice monitoring were carried out by NASDA in 1985 and in 1988 in the Okhotsk Sea off the north coast of Hokkaido, Japan. All the sea ice within this region are one year or younger ice. This situation simplifies the evaluation of MSR. The MSR and VTIR equivalent in performance to MOS-1 sensors were mounted on an airplane, turbo-prop type "Merlin IV A", and several flights were conducted. The parameters of the data used in this study are shown on Table 1.

Table 1. Parameters of the analyzed data.

| Date | Flight Altitude | Sensor | Swath Width | Band No. | Frequency / Integration time (Wave Length) | Resolution |
|-----------------|-----------------|---------------|-------------|----------|--|------------|
| Jan. 25 1985 | 500 m | Airborne MSR | 200 m | 3 | 31 GHz / 10 msec | 10 m |
| | | Airborne VTIR | 1340 m | 1 | 0.5 ~ 0.7 μ m | 2.4 m |
| Feb. 9 1988 | 5 Km | Airborne MSR | 1.8 Km | 3 | 31 GHz / 10 msec | 114 m |
| | | Airborne VTIR | 8.4 Km | 1 | 0.5 ~ 0.7 μ m | 15 m |
| | 900 Km | MOS-1 MSR | 320 Km | 3 | 31 GHz / 10 msec | 23 Km |
| | | MOS-1 VTIR | 1500 Km | 1 | 0.5 ~ 0.7 μ m | 0.9 Km |

2.1 Ice thickness

Firstly, the basic ability of MSR for detecting ice thickness difference was evaluated with the low altitude (800m) airborne MSR and VTIR data of the Okhotsk Sea for Jan. 25, 1985. The Fig.1 shows an example of MSR image. The image is underscanned because of the low flight altitude. The footprint of MSR are overlaid on the VTIR image as shown on Fig. 2. By comparing the image with simultaneously taken aerial photos, 7 plane ice zones were selected as training area. Open water were not observed in this site. Fig 3. shows a linear relationship between the VTIR data value and MSR brightness temperature. If we ignore the effect of the snow cover, the VTIR brightness value of ice increases with the thickness of ice. The result suggests the basic ability of MSR for ice thickness difference detection.

2.2 Ice Concentration

Sea ice concentration information are significant for routing ships safely through ice. Originally, sea ice charts were made manually with various information mainly corrected with visual observation from shore, from ships, from low altitude airplane. Nowadays, remotely sensed images, such as NOAA AVHRR images, are becoming utilized for making sea ice charts. Though the visual definition of sea ice concentration using remotely sensed image may change with the correlation of the size of the ice floes and the resolution of the image, if we assume a linear relationship between the brightness value of the remotely sensed data and ice concentration, the ice concentration can be estimated by the following formula. (after Zwally et al, 1982, Alexandrov et al, 1987)

$$I_{cz} = \sum_{i \in z} \{ (D_i - D_w) / (D_1 - D_w) \} / \sum_{i \in z} 1 \quad (1)$$

where:

I_{cz} = Ice concentration of zone Z

D_i = Brightness value of the pixel i

D_w = Brightness value of open water

D_1 = Brightness value of white ice

It should be noted that this formula calculates "average ice concentration" of the specified ice zone Z. If one wants to correlate the calculated ice concentration with visually identified conventional ice concentration, one should select ice zones where the sea ice are evenly distributed within the zone.

Fig 4. shows an airborne MSR image of the Okhotsk Sea for Feb. 9, 1988. The altitude of the flight was 5000m and the resolution of the MSR data for 31GHz is 114m. In the test site, total of 18 ice zones for various ice concentration were selected manually as shown on Fig. 5. The ice concentration of each zone were calculated from VTIR visible data using formula(1). The relationship between the MSR brightness temperature and

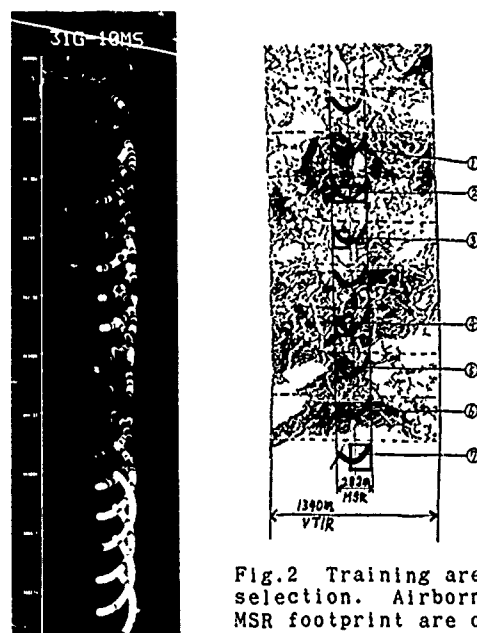


Fig.1. An example of airborne MSR image. (31GHz)

Fig.2 Training area selection. Airborne MSR footprint are overlaid on airborne VTIR image and 7 training area were selected.

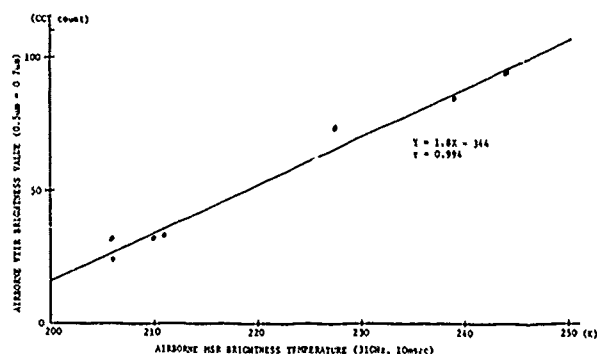


Fig. 3 Relation between MSR brightness temperature and VTIR brightness value which reflects relative ice thickness.

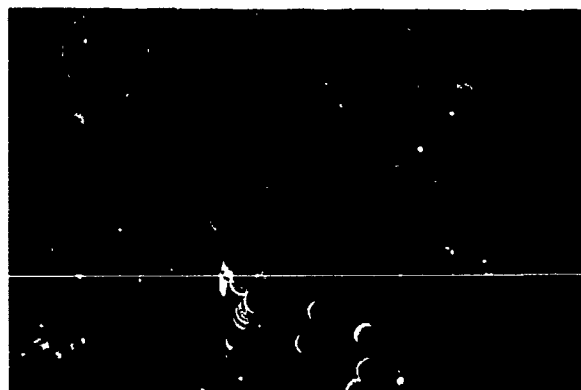


Fig. 4 Airborne MSR image of the Okhotsk Sea for Feb. 9, 1988. (31GHz)

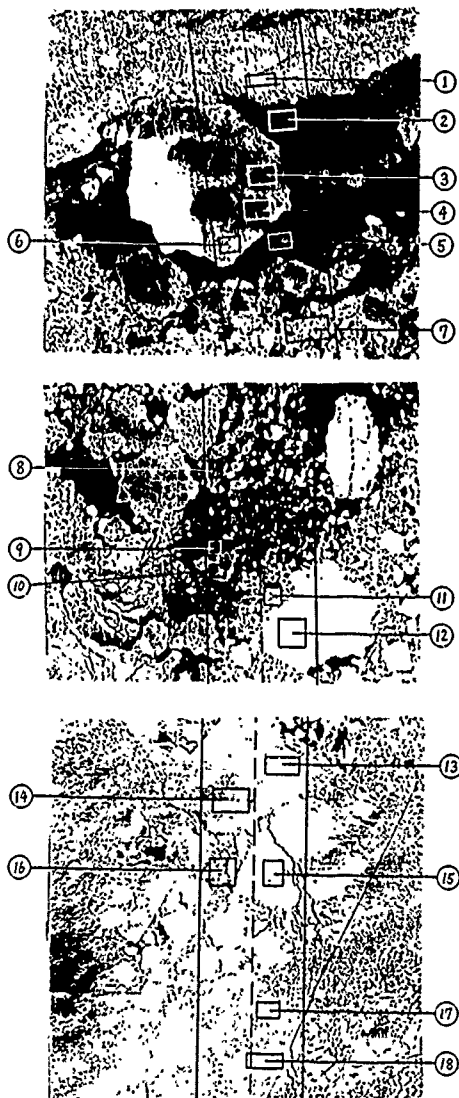


Fig.5 Training area overlaid on aerial photo.

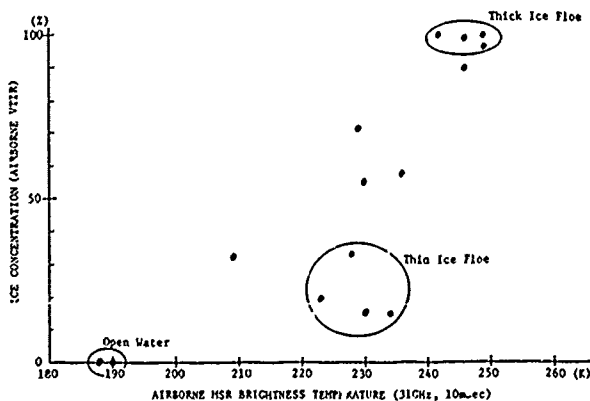


Fig. 6 Relation between airborne MSR brightness temperature and ice concentration derived from VTIR brightness value.

ice concentration derived from VTIR data is shown on Fig. 6. A obvious brightness temperature gap (30K difference) between the open water and ice floes was observed. The tendency of the increase of MSR brightness temperature with the increase of ice concentration was slightly observed. Though thin ice floes and ice-water mixed area showed clear difference in VTIR derived ice concentration, the airborne MSR brightness temperature of thin ice floe are mostly overlapped with that of ice-water mixed area.

3. Satellite Data Analysis

Based on the result of the airborne data analysis, the satellite data are analyzed. The parameters of the data used for the analysis are shown on Table 1.

Fig. 7 and Fig. 8 show the MSR image and VTIR image respectively of the Okhotsk Sea for Feb. 9, 1988. Total of 15 training area with different ice concentration were selected manually with the simultaneously corrected VTIR image as shown on Fig. 9.



Fig. 7 MOS-1 MSR image of the Okhotsk Sea for Feb. 9, 1988. (31GHz)

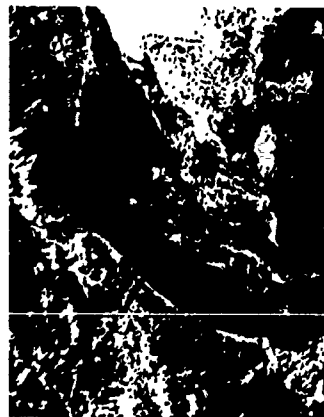


Fig. 8 MOS-1 VTIR visible image of the Okhotsk Sea for Feb. 9, 1988.

The ice concentration of each training area were calculated from VTIR visible data using formula(1). Fig. 10 shows the relationship between the MSR brightness temperature and the ice concentration derived from the VTIR data. The relation between brightness temperature and ice concentration became more linear compared with the airborne data. This tendency can be explained as follows. Compared with airborne MSR, MOS-1 MSR has much lower resolution. It is obvious that thin ice floes with the size of MOS-1 MSR footprint (23Km) can hardly exist. Accordingly, intermediate brightness temperature of MOS-1 MSR may represent the grade of ice concentration. This result suggests the capability of estimating ice concentration from MOS-1 MSR data.

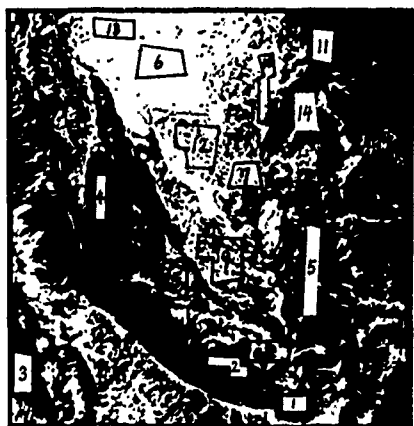


Fig. 9 Training area overlaid on MOS-1 VTIR visible image of the Okhotsk Sea for Feb. 9, 1988.

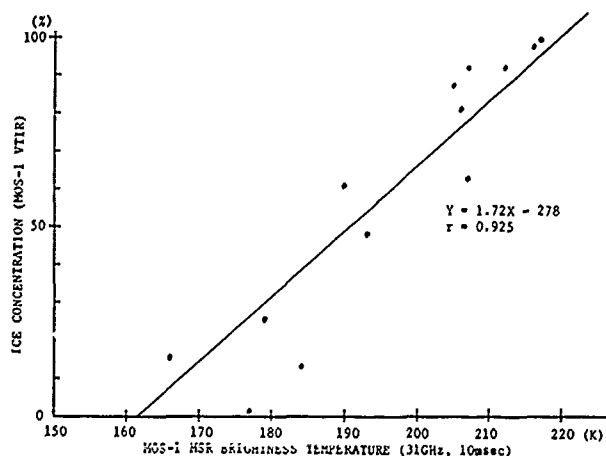


Fig. 10 Relation between MOS-1 MSR brightness temperature and ice concentration derived from MOS-1 VTIR brightness value. The gain and offset may change scene by scene.

4. Conclusion

The airborne and the MOS-1 MSR data were analyzed to evaluate the capability of MSR for sea ice monitoring. Ice thickness and ice concentration of the test sites were estimated from the simultaneously corrected VTIR data and were compared with MSR data. The airborne MSR brightness temperature showed a good linear relationship with sea ice thickness. More than 30K brightness temperature difference was observed between the open water and ice floes. The tendency of the increase of ice concentration with the increase of MSR brightness temperature is slightly observed. As for the MOS-1 MSR data, due to the large footprint size of the MSR, the MSR brightness temperature showed a linear relation with the sea ice concentration estimated by the VTIR. Considering the observation ability of MOS-1 MSR with wide coverage, through cloud, and both day and night, the usefulness of MOS-1 MSR for sea ice monitoring is promising.

5. Reference

1. Alexandrov V.Y., Bushuev A.V., and Loshchilov V.S., "Sea Ice Concentration Assessed from Aerospace Imagery", *Sov. J. Remote Sensing*, Vol.5(2), pp181-189, 1987
2. Steffen K., Barry R., Schwegler A., "DMSP-SSM/I NASA Algorithm Validation using Primarily Landsat, and Secondarily DMSP and/or AVHRR Visible and Thermal Infrared Satellite Imagery", NASA-CR-182979, 1988
3. Comiso J.C., Zwally H.J., "Antarctic sea ice concentration inferred from Nimbus 5 ESMR and Landsat imagery", *J. Geophys. Res.*, 87(C8), pp5836-5844, 1982

AN EVALUATION OF MOS-1 MESSR DATA FOR LAND APPLICATIONS

W. Murray Strome
PCI Inc.
50 West Wilmot St., Unit 1
RICHMOND HILL, Ontario, Canada
L4B 1M5
Phone: (416) 764-0614
Fax: (416) 764-9604

Ellsworth F. LeDrew
University of Waterloo
WATERLOO, Ontario, Canada
N2L 3G1

ABSTRACT

The data from the MOS-1 satellite MESSR sensor were processed by the Canada Centre for Remote Sensing (CCRS) for three scenes in Southern Ontario which are being evaluated by the authors for several land applications, including land use, agriculture and forestry. The image data are geometrically corrected to map coordinates for ease of comparison and combination with the large amounts of data from other sources using the PCI EASI/PACE image processing and analysis system. Using the same system, enhancements, classification and other analysis techniques are applied.

The MESSR data are examined for their independent usefulness, as well as for the additional information and utility of the data when combined with information obtained from other satellite and airborne sensors as well as ground observations.

THE DESIGN AND IMPLEMENTATION OF THE EUROPEAN RADAR CROSS SECTION DATABASE (EURACS)

A L Wielogorski¹, D Brewster², I R Barnes¹, P N Churchill³, N J Veck², G C Deane¹ & A Sieber³

¹
Hunting Technical Services Ltd
Thamesfield House
Boundary Way
Hemel Hempstead
Herts, HP2 7SR, UK

²
GEC-Marconi Research Centre
West Hanningfield Road
Great Baddow
Chelmsford
Essex, CM2 8HN, UK

³
Institute for Remote Sensing Applications
Commission of the European Communities
Joint Research Centre
Ispra Establishment
21020 Ispra (VA) Italy

ABSTRACT

The Euracs database has been set up in response to a need within Europe to rationalize research into microwave remote sensing. This database is designed to store the radar cross section and associated data of ground cover at microwave frequencies. Its development and installation have been commissioned by the Institute of Remote Sensing Applications at the Joint Research Centre (JRC), Ispra Establishment of the Commission for European Communities and undertaken by Hunting Technical Services and GEC-Marconi Research Centre.

The database has been developed as a facility for use by researchers into microwave remote sensing, initially to assist research into a range of activities including backscatter modelling, algorithm development and microwave sensor design. In the long term improved modelling techniques will be employed within the database to enhance the efficiency of data analysis. This will encourage other applications such as algorithm development for microwave product simulation. In addition, the database can be used to assist in campaign planning and training in image processing.

1. INTRODUCTION

The European radar cross section database (Euracs) is a central store of calibrated and validated radar cross section (rcs) values and associated data. It covers a full range of microwave sensors and platforms. Euracs is a data analysis tool permitting access to both accumulated data and a range of applications such as backscatter modelling, sensor design, algorithm development and interpretation. The background to the development of Euracs has previously been described by Churchill and Sieber (1988), however, it is worthwhile reiterating the overall objectives. The database must be fully quality checked, it must be flexible with defined limits of validity; all parameters present must be defined. The database must be able to recreate experiments from sensor to ground parameters, this implies the ability to hold data from various sources. Finally the database must be fast and easily available.

Euracs has been developed in response to a need to rationalise European research into microwave remote sensing. It will summarise the current state of empirical knowledge and facilitate access to accumulated data. As a result of the introduction of approved sampling and ground collection methodologies, as well as rcs measurement methodologies, Euracs will lead to a uniform approach to microwave remote sensing programmes.

Four types of potential Euracs user have been identified and Euracs has been designed to satisfy each of their requirements. The rcs modeller needs to evaluate the average parameter value of each field and the variability present both within and between fields. Basic questions, such as how rcs values change with radar sensor parameters and crop parameters, will be answered empirically. The sensor engineer will be

able to extract values for particular ground cover types and thus design a sensor most suited to the desired application. The algorithm developer may utilise the application demonstrators, or backscatter models, within the database to facilitate development of analysis techniques. Finally the campaign planner can access information about existing campaigns and methodologies employed.

2. DATABASE PARAMETERS

The development of Euracs, as originally conceived by JRC, has taken about 3 years. The programme of development began with the European Agrisar 86 (JRC, 1986), Agriscatt 87 (HTS, 1988) and Agriscatt 88 airborne experiments, in which rcs data were collected in a coordinated manner using standard data collection methodologies. These experiments identified clear definitions for the parameters required, the data collection and processing methodologies, the quality assurance procedures and data integrity.

The data currently held within Euracs are mainly based on information collected in the Agriscatt 87 campaign. However additional tables have been constructed to enable the future inclusion of multi-polarization, polarimetric, multi frequency and laboratory data and data from other sensors.

The data collected and stored within Euracs can be grouped into parameters describing the target, sigma values (rcs), environment, data quality, instruments and administrative information. These parameters or attributes are grouped together to form entities or tables.

2.1 Target parameters

The defined target parameters (such as species, crop height and phenology) are based on those used in the Agriscatt 87 dataset (HTS, 1988), but equally applicable to a range of spaceborne, airborne or lab-based experiments. Data were collected, using these definitions over six sorties in 1987. However not all crops are of the same physiology, and so not all the specified parameters can be collected for every crop. In such cases the position in the relevant table is filled with a NULL, this implies to the database that there is no measurement. Subsequently when using Euracs these values are non-existent, and cannot be accessed in any operation.

The database logical design of target parameters is such that the most frequently changing, ie 'unstable', parameters are contained within the table 'CROP_DES'; whereas parameters that may remain 'stable' throughout the year are contained in a higher order entity or table such as 'CROPSUR'. Such separation of the data into tables avoids the need for duplicating data at different levels.

2.2 Sigma Parameters

For in-situ crop targets each field is notionally divided into

subfields. The objective of this is to test for homogeneity of the rcs values across a field. The subfields are assumed to lie at a minimum distance from the field edge. This minimum distance would typically be given by $10 + 2R$ meters, where R is equal to the minimum of the range and azimuth resolution.

Database records for the entire field only contain subfields with 'good' data. For example with a goodness of fit below a 5% threshold; a field yielding such results is judged as being homogeneous.

There is also a Euracs entity for laboratory measurements of rcs targets. For multi-polarisation data there is an entity which enables the storage of full Muller matrices, mean and standard deviation and goodness of fit. An entity has also been designed for accepting ERS-1 and other satellite-borne data.

2.3 Environment Parameters

These are contained within a single entity which records precipitation, cloud cover, sunlight and wind parameters.

2.4 Quality Parameters

For each measurement there are stored associated quality parameters detailing, for example, spatial and radiometric resolution, the point spread function, antenna pointing accuracy, signal/noise ratio and calibration approval for each instrument. These parameters enable the user to determine the quality of the output, both in terms of the information obtained and to ensure that the output is of a specified quality.

2.5 Instrument Parameters

The parameters stored in these associated entities are concerned with the instruments and platforms deployed. The main recipient of this information is expected to be the sensor engineer. There is also an associated entity PROCESSOR which gives the processing algorithms used to convert raw data to rcs values.

2.6 Administrative parameters

The parameters associated with these entities give details about which organizations collected ground data and rcs data, over various sites. They also give details of the sponsoring agencies. These data are mainly of interest to campaign planners.

3. DATABASE MANAGEMENT SYSTEM AND THE LOGICAL DESIGN

The data contained within Euracs are managed using a Relational Database Management System (RDBMS). The software package chosen to perform this task was supplied by Sybase Software Ltd. Euracs has been developed on a SUN 4 Microsystems computer running the UNIX operating system.

There are numerous advantages to using the relational model for data storage. The software allows easy access to the data by means of the SQL (Structured Query Language) interface, or by user friendly forms that are interactively filled in on the computer display screen. The system also enables the use of data via networks. A further advantage is the ability to have automatic built-in checks on the integrity and quality of the data.

The database consists of a number of tables or entities each with their own list of attributes or headings. The choice of parameters (attributes) associated with each entity has already been discussed. The design of the structure of these tables has been made to satisfy the criteria of the third normal form (Hove). The advantages of this structure are that it avoids data redundancy and the design can readily be expanded to accept new attributes and new tables.

In Figure 1, The Logical Data Model, the entities or tables and possible relationships generated for the data within Euracs are indicated. The relationships in the logical model are shown in a top down format. Along the top, of this diagram the highest order entities are shown. These contain attributes common to all entities directly below them. A table with

a double arrow pointing to it is considered to be below the table from which the line emanates. A line with single arrows shows equivalent associated data. Therefore, each parameter has a primary table, i.e. that in which it first appears. Any information contained in a higher order entity must, by definition, be common to the entities below it, hence avoiding the need for data repetition. This structure also maintains data integrity, in that any row within an entity, in order not to be removed from Euracs, must have all the associated rows above it also present within the database.

Each table within Euracs has a set of identifying attributes (parameters) with the following properties:- (a) each non-identifying attribute is dependent on all identifying attributes and (b) each non-identifying attribute is not dependent on another non-identifying attribute. The identifying attributes or parameters uniquely identify the row within the entity. To achieve this existing tables are divided into new tables and attributes are moved between tables.

4. DATA QUALITY AND INTEGRITY

Data integrity is maintained by having a three level table structure to the database. The primary tables are the 'real' quality checked and assured tables. These tables contain parameters which have all their associated parameters present in associated tables. Each rcs value has its associated target parameters contained in the database.

The second primary table type is the intermediate table. An intermediate table is one whose quality and value is assured. However, associated parameters might not yet be present, thus the measurement is not valid for all possible applications. Both these tables can be accessed by a user and the user will be aware of which type of table he is obtaining data from.

There also exist temporary tables. These are invisible to the user and contain data which have failed quality assurance measures in some respect. These data are placed in a temporary table for further investigation. From the temporary table the parameter values are either rejected or corrected and placed into intermediate tables, pending dataset completion.

As a result of the presence of an entity relationship diagram within Euracs, data integrity is further maintained; because, no row can remain in the real tables unless all higher order rows are also present. This 'cascade delete' runs constantly. It also ensures data are entered in a logical manner.

Data input to the database are subject to various quality checks. Data are automatically checked to ensure that each parameter falls within a permitted range of values or, if of a character datatype, that it is one of the permitted values. The automatic check also compares the units of measurement used and converts them if necessary to a Euracs standard. Other automatic checks will include a double row check, to ensure rows of data in large files are not duplicated.

It is also envisaged in the future that automatic relational checks will be performed. These may consider relations not only within tables but also between tables. Relations such as permissible crop height or phenology, given sowing date and weather conditions, or field reference given sitename will be considered.

There is also a manual data integrity check, which is a check that measurement methodologies are consistent with Euracs standards. This also includes interpolation methodologies and parameter compatibility. This check will also decide on the logical definition of any new parameters or entities that may be added to the database.

5. USING EURACS

The data obtained from Euracs are in an alphanumeric form, with associated summary statistics if required. It is intended that at a later date graphics and statistics packages will be added, enabling three types of user output. Euracs output is given as rows of data in specified columns derived from user specified tables, joined in a user specified way with user specified conditions attached. Euracs is designed as a read only database for users.

FIGURE 1: THE LOGICAL DATA MODEL



FIGURE 2: FORMS INTERFACE

RCSCROPALL

RUN! CLEAR!

region: country: species: winter wheat
 sitename: variety:
 fieldref: from_month: Jun
 to_month: Jun

freqband: 1 inc_angle_min: 20
 frequency: inc_angle_max: 35
 polarization: ☒ ☐ ☐

number of rows found = 74

| frequency | polarization | incidence_angle | rsc_mean | rsc_std_dev |
|-----------|--------------|-----------------|----------|-------------|
| 1.2 | hh | 30 | -14.26 | 0.0028 |
| 1.2 | vv | 23 | -15.97 | 0.0056 |
| 1.2 | vv | 20 | -14.45 | 0.0056 |
| 1.2 | hh | 20 | -11.0 | 0.0048 |
| 1.2 | hh | 31 | -13.04 | 0.0032 |
| 1.2 | vv | 22 | -15.88 | 0.0017 |

5.1 User interface

The RDBMS provides several user interfaces. The most basic of these is the Structured Query Language (SQL) which can directly access the database. This method of query is extremely flexible. It does, however, require user knowledge of SQL.

A less flexible but more user friendly query interface is by what is termed 'stored procedure'. Stored procedures contain SQL statements but they have unspecified values for the parameters in the conditions clause. Input parameters have been declared in the procedure as variables that can be specified by the user as required. If they are not specified then a wildcard is placed in the SQL statement and no restrictions are applied. It is also possible to join tables to themselves and thus to compare, for example, horizontal and vertical polarisations for the same conditions.

The RDBMS also provides an interface based on forms. The database presents forms on which the user fills in the 'blanks'. The form contains predetermined parameters as indicated in the sample form (Figure 2). The form is more user friendly than a stored procedure as it also has a help facility for filling out the form. As parameters are filled in, the form logic will eliminate impossible parameters. There exists a forms hierarchy whereby supplementary forms can be called upon and further limiting parameters entered.

5.2 Programming Language Interface

Modelling backscatter requires access to the database from a language such as C or Fortran. An application demonstrator has been written in C; this is a theoretical model for the variation of rcs with crop height and incidence angle. This model can be used to predict rcs values with different conditions. The demonstrator is designed to show how data in the database can be accessed in a useful manner, to solve a real problem. The interface to C can also be used to build other data models.

5.3 Networking

Euracs can be accessed by both local and remote links. It is successfully accessed by a Vax. Once a user has logged into the SUN by whichever link, he can use any software package present in the SUN. The SQL and stored procedure statements have successfully been sent to Euracs using an X25 modem over the public packet switched data network. This was achieved by logging into the SUN running the DBMS remotely from an IBM PC/AT. The only condition for the communication software and terminal is that it can emulate a VT100 terminal. It is also envisaged that users will be able to access the database using their own compatible dbms software.

6. FUTURE WORK

At present the database contains about 6Mb of data from the AGRISCATT 87 campaign; this represents rcs values collected using a scatterometer over 61 fields in 5 European sites, together with associated ground data. In the future Euracs will hold multi-polarization SAR data and laboratory data. Improvements in quality checking and the user interface will also be undertaken. It is also intended to expand the scope of Euracs by including data from different target types, such as forests, ice and snow.

7. SUMMARY

Euracs has been implemented at JRC, the validation phase is now in progress and Euracs will be available to users in the near future on a read only basis; together with full user documentation, although Euracs will be constantly updated and extended.

Database tables have been devised that provide the framework for accepting rcs data from a number of different instrument types. Quality procedures have been constructed to ensure that rcs data entered are correctly calibrated and that ground data are collected to a uniform standard. Columns can also be added to the tables if additional parameters are to be stored.

User access is designed to be flexible and friendly, either on line from the screen or through a computer program. Data integrity measures are such that a user can have a 100 per cent guarantee of valid data from real tables, and at worst quality assured data from intermediate tables.

The Euracs database will provide a valuable source of data for users wishing to understand the interaction of microwaves with ground cover.

REFERENCES

Churchill P & Sieber A, "Euracs: A European Radar Cross Section Database", in Remote Sensing: Moving towards the 21st Century, Proceedings IGARSS'88, ESA SP-284, 1988

Howe D R, "Data Analysis for Database Design" Edward Arnold

Hunting Technical Services, "Agriscatt 1987 Campaign: Coordination of Ground Data- Final Report", for JRC Contract No. 3171-87-06 EO-ISP-GB, 1988

Joint Research Centre, Ispra Establishment "Agrisar 86: Preliminary Investigators' Report" S.P./1.87.25 1987

DIGITAL SIMULATION OF TERRAIN BACKSCATTERING CROSS SECTIONS

Albert W. Biggs

Department of Electrical and Computer Engineering
University of Alabama in Huntsville
Huntsville, AL 35899, USA

Abstract - Digital simulation of the terrain seen by airborne imaging radars is presented as a method of evaluating these imaging radars during simulated flights. Terrain simulated includes vegetation, oceans, Arctic sea-ice, geological and man-made structures, fresh water bodies, and glacial snow or ice. Each category has a target signature given by its radar backscatter cross section.

Computer programs of backscatter cross sections are found in a terrain reference file. Actual segments of terrain are simulated by reference to the file in the radar computer data bank. During the flight, the radar compares its reception with terrain signatures in its data bank like a road map in a car.

Key Words - Backscatter, Terrain, Simulation, Digital

INTRODUCTION

One area of radar simulation studies is the digital simulation of the terrain seen by a missile or aircraft radar during its simulated flight. Terrain simulated includes vegetation, sea-ice, glacial ice, geological structures, man-made structures, and fresh water bodies. Each terrain category is described by its radar backscatter cross section, which has a target signature associated with that category. Computer programs of backscatter cross sections are formed in a reference file so that an actual segment of terrain can be simulated by reference to the terrain reference file. A related application is the inclusion of a terrain reference file in an aircraft or missile computer data bank. During the flight phase, the aircraft or missile can compare its radar reception with the terrain signatures in its computer data bank similar to using a road map on an automobile trip.

BACKSCATTERING CROSS SECTIONS OF TERRAIN

Terrain backscattering cross sections appear in CAPITAL LETTERS. Variables appear within (PARENTHESES). PROGRAM SIGTERRA is a library for terrain backscattering cross sections or coefficients [1] of different categories of terrain such as vegetation, oceans, Arctic sea-ice, man-made structures, geological structures, fresh water bodies, and glacial snow or ice. Backscatter coefficients (SIGTV) vary with the angle of incidence (THETA) of the radar signal and the terrain, frequency (FREQG), polarization of the backscattered signal with respect to the source signal (POLAR), and the terrain itself (TERRA).

SUBROUTINE OCEAN computes backscatter coefficients for sea states (STA) and wind directions on oceans. Sea states [2] indicate ocean wave roughness. A calm, mirror-like surface with no waves is sea state 'one.' As ocean waves begin to develop from winds and tidal actions, sea states increase to higher numbers. Sea state 'seven' is a very rough surface, where high waves are developed from high wind speeds. Storms and wind directions can be observed with the radar backscatter from ocean waves. High rough waves have larger backscatter coefficients, while low gentle waves have very small backscatter coefficients. Wind direction for a given wind speed is determined from the different radar cross section seen when it looks downwind, crosswind, or upwind at the waves (SEA).

SUBROUTINE LAKES computes backscatter coefficients for fresh water bodies (WATER). This terrain includes lakes, rivers, canals, reservoirs, marshes, and swampland. Large lakes and the Great Lakes (Lakes Erie, Michigan, Huron, Ontario, and Huron) have high waves similar to and sometimes higher than those on oceans during storms. Sea state roughness levels (STA) during storms over oceans are often exceeded by wave heights found in intense storms in the Great Lakes. The scattering coefficients for these water bodies may be larger than those for oceans. The remaining water bodies have relatively smooth surfaces, so that radar images indicate negligible surface roughness. The backscattered signals for these surfaces are predominantly specular, following Snell's Law of Reflection.

Related to ocean and fresh water body terrain are Arctic sea-ice and fresh water lake ice. SUBROUTINE SALICE includes scattering coefficients from different forms of Arctic sea-ice (ICE). This ice is known as 'Arctic sea' ice because of its formation and presence in the Arctic Ocean. Sea-ice is encountered in other salt water bodies where conditions are similar to those in the Arctic Ocean. From a remote sensing aspect, much radar and IR imaging data [3]-[4] were collected in the Arctic Ocean because of economic and strategic interests. When sea-ice is less than a year old, it is known as 'winter ice.' When it is more than a year old, it becomes 'old ice' or 'polar ice.' Polar ice is rougher than winter ice because it has broken several times, piled in ridges, and joins other polar ice in 'pack ice.' Polar ice also has a lower brine, or salt water content than winter ice, and the distribution of brine in brine pockets introduces internal scattering. This brine pocket distribution and the surface roughness of sea-ice alters the scat-

tering coefficients so that sea-ice forms and depths may be identified [5].

Fresh water ice forms are sound in SUBROUTINE SNOW. Fresh water ice (GLA) is formed in precipitation as snow, hail, and sleet. It is formed in glaciers, snowfields, fallen snow, and icebergs. In glacial ice and in snowfields, internal scattering from dielectric ice spheres [6] may actually indicate much larger scattering coefficients than those due to surface roughness alone. This phenomena occurs because of focusing of incident waves upon the dielectric spheres. Although fresh water lake ice is formed in the water, it is included in this terrain category because of its physical features. During the spring, lake ice breaks up and form ridges similar to those of sea-ice. Radar images have been made in the Great Lakes and in Finnish rivers [7] for ice distribution. Imaging studies [5] enables ice breakers to keep the Great Lakes open for longer periods during the year. Icebergs are a form of fresh water ice because of their formation in coastal terminating glaciers. Detection of icebergs is an important application in remote sensing over the North Atlantic Ocean.

Scattering coefficients for geological structures appear in SUBROUTINE ROCKS. These terrain classes [8] includes variations in geological structural features. It includes bare soil, sand, lava beds, hills, mountains, valleys, rocks, boulders, tundra, permafrost, clay, and similar terrain (GEO).

SUBROUTINE VEGET describes all forms of vegetation. It includes scattering coefficients for prairies, grasslands, bushes, shrubs or flower cover, tree cover or forest, and agricultural crops (VEG). Further categorization introduces orchards, deciduous trees, and evergreen trees. Agricultural crops have categories based on plant size, leaf size and distribution [9], stem and branch shapes [10], moisture content and its distribution in vegetation, and fruit characteristics. In addition to apples, oranges, and berries, fruit includes pea pods, cucumbers, grains of wheat, and similar varieties. "Radar signatures" of agricultural crops identify different plants in the form of radar signatures [11]-[12] with frequency, polarization (POLAR), and angle of incidence (THETA).

The last category is all man-made terrain in SUBROUTINE STRUCT. Some of the man-made structures are houses, buildings, roads or highways, railroad tracks, airfields or runways, aircraft hangars or farm buildings, railroad tracks, telephone lines, power lines, ships, boats, barges, bridges, automobiles, trucks, buses, and trains.

PROGRAM SIGTERRA, file name SIGMATERRAIN, receives input data (TERRA) to indicate vegetation with '1', oceans with '2', sea-ice with '3', glacial snow with '4', geological structures with '5', man-made structures with '6', and fresh water bodies with '7'. After the terrain "number" is read as TERRA, the vegetation category (VEGETA), the radar direction with respect to wind direction (radar looks to wind direction and at the ocean downwind, upwind, or crosswind) and sea state (OCEANS), form of sea ice (SEAICE), form of glacial snow or ice (GLASNO), geological structure (GEOLOG), man-made structures (MANMAD), or fresh water bodies (WATER) are read in character format. The frequency of the incident wave (FREQG) in Gigahertz (GHz) is then read. Polarization (POLAR) is next read as '11' when both incident and backscattered waves are horizontally polarized, or HH. It is '22' when both waves are vertically polarized, or VV. '12' describes

horizontally polarized incident waves and vertically polarized backscattered waves, or HV. '21' describes vertically polarized incident waves and horizontally polarized backscattered waves, or VH. Right circular polarization of the incident wave, or RC, is '10', while left circular polarization of the incident wave, or LC, is '20'. The last variable read is the angle of incidence (THETA) of the radar wave as it reaches the terrain (in degrees).

Based on TERRA, a terrain subroutine is called. As described above, it will be VEGET, OCEAN, SALICE, SNOW, STRUCT, or LAKES. This subroutine opens a file (TERRA.DAT) to compute the scattering coefficient for a given frequency (FREQG), angle of incidence (THETA), polarization (POLAR), and terrain. It also calls SUBROUTINE INTERP. These variables are obtained from experimental measurements and analytical models.

INTERP computes the backscattering coefficient (SIGFTV) by interpolation between frequencies (FRINT(I)) and incremental angles of incidence (J). Sets of backscattering coefficient versus theta curves are computed for frequencies of 5, 10, 15, 20, 25, and 30 GHz. Theta increments are 0, 5, 10, 15, 20, and 25 degrees. The interpolation between theta increments is initially made to obtain SIGTV(I,POLAR,TERRA) and SIGTV(I+1,POLAR,TERRA). The second interpolation between frequency increments is then made to obtain SIGFTSV(POLAR,TERRA).

Examples of backscattering cross section computations appear in Fig. 1, backscattering coefficients for sea state five and radar looking downwind, in Fig. 2, backscattering coefficients for sea state seven and radar looking crosswind, and in Fig. 3, backscattering coefficients for vegetation in the form of wheat-fields.

Several sources are [12]-[14] used for scattering coefficient (SIGMA) versus angle of incidence (THETA) curves. In addition to "Terrain Scattering Properties for Sensor System Design" [12], there are other excellent handbooks [15] and references on remote sensing [16]-[18].

REFERENCES

- [1] Biggs, A.W., "Radar and Infrared Remote Sensing of Terrain, Water Resources, Arctic Sea Ice, and Agriculture," 32nd Symposium of EWPP/AGARD, Oberammergau, Federal Republic of Germany, May 24-28, 1983.
- [2] Chaudry, A.H. et al, "Tower-Based Broadband Backscattering Measurements from the Ocean Surface in the North Sea," Proceedings of IGARSS' 86 Symposium, Zurich, Sept. 8-11, 1986.
- [3] Onstott, R., "An Inter-Sensor Comparison of the Microwave Signatures of Arctic Sea Ice," Proceedings of IGARSS' 86 Symposium, Zurich, Switzerland, Sept. 8-11, 1986.
- [4] Onstott, R., "Theoretical and Experimental Study of the Radar Backscatter of Arctic Sea Ice," Proceedings of IGARSS' 87 Symposium, University of Michigan, Ann Arbor, May 18-21, 1987.
- [5] Biggs, A.W., "Sea Ice Thickness Measurements with Short Pulse Radar Systems," URSI Symposium on Propagation in Non Ionized Media, Le Baule, France, April 28- May 6, 1977.

- [6] Biggs, A.W., 'Volume Scattering from Ice and Water in Inhomogeneous Terrain,' 20th Symposium of EWPP/AGARD, The Hague, Netherlands, March 25-29, 1974.
- [7] Toikka, M., 'The use of Radars to Measure the Distribution of Ice and Frazil in Rivers,' Proceedings of IGARSS' 87 Symposium, University of Michigan, Ann Arbor, May 18-21, 1987.
- [8] MacDonald, H.C., 'Techniques and Applications of Imaging Radars,' in Siegal, B.S. and A.R. Gillespie, editors, Remote Sensing in Geology, John Wiley & Sons, New York, pp. 296-326.
- [9] Fung, A.K. et al., 'Scattering Models for Some Vegetation Samples,' Proceedings of IGARSS' 87 Symposium, University of Michigan, Ann Arbor, May 18-21, 1987.
- [10] Karam, M.A. and A.K. Fung, 'A Scattering Model for Defoliated Vegetation,' Proceedings of IGARSS' 86 Symposium, Zurich, Switzerland, Sept. 8-11, 1986.
- [11] Biggs, A.W., 'A Review of Papers Presented at the URSI Commission F Symposium on Microwave Signatures in Toulouse, France, 35th Symposium of EWPP/AGARD, NATO, on Les Signatures d'Objets, London, England, October 8-13, 1984.
- [12] Cosgriff, R.L., W.H. Peake, and R.C. Taylor, Terrain Scattering Properties for Sensor System Design (Terrain Handbook II), Report No. 181, Ohio State University, Columbus, Ohio, 1960.
- [13] Peake, W.H. and T.L. Oliver, 'The Response of Terrestrial Surfaces at Microwave Frequencies,' TR ALAL-JR-70-301, AF Avionics Lab, Wright Patterson AFB, Ohio, 1971.
- [14] Moore, R.K., 'Radar Imaging Applications; Past, Present and Future,' in Biggs, A.W., editor, Proceedings of AGARD Conference on Propagation Limitations in Remote Sensing, USAF Academy, CO, June 21-25, 1971.
- [15] Ruck, G.R., D.E. Barrick, W.D. Stuart, and C.K. Krichbaum, Radar Cross Section Handbook, vol. I, Plenum Press, New York, New York, 1970.
- [16] Ulaby, F.T., R.K. Moore, and A.K. Fung, Microwave Remote Sensing: Active and Passive, vol. I--Microwave Remote Sensing and Radiometry, Addison Wesley Publishing Co., Reading, Massachusetts, 1981.
- [17] Ulaby, F.T., R.K. Moore, and A.K. Fung, Microwave Remote Sensing: Active and Passive, vol. II--Radar Remote Sensing and Surface Scattering and Emission Theory, Addison Wesley Publishing Co., Reading, Massachusetts, 1982.
- [18] Ulaby, F.T., R.K. Moore, and A.K. Fung, Microwave Remote Sensing: Active and Passive, vol. III--From Theory to Applications, Artech House, Inc., Dedham, Massachusetts, 1986.

BACKSCATTERING COEFFICIENT FOR DIFFERENT SEA STATES AND LOOKING DOWNWIND, OR

RADAR LOOKS IN THE DOWNWIND DIRECTION WITH SEA STATE FIVE

BACKSCATTERING CROSS SECTION SIGMA VERSUS THETA FOR SEA STATE FIVE

| FREQUENCY | ANGLE OF INCIDENCE THETA IN DEGREES | | | | | | |
|-----------|-------------------------------------|-------|--------|--------|--------|--------|--|
| IN GHZ | 0 DEG | 5 DEG | 10 DEG | 15 DEG | 20 DEG | 25 DEG | |
| 5.00 | 20.00 | 17.00 | 14.00 | 12.00 | 10.00 | 8.00 | |
| 10.00 | 17.00 | 14.00 | 11.00 | 9.00 | 7.00 | 5.00 | |
| 15.00 | 14.00 | 11.00 | 8.00 | 6.00 | 4.00 | 2.00 | |
| 20.00 | 11.00 | 8.00 | 5.00 | 3.00 | 1.00 | -1.00 | |
| 25.00 | 8.00 | 5.00 | 2.00 | 0.00 | -2.00 | -4.00 | |
| 30.00 | 5.00 | 2.00 | -1.00 | -3.00 | -5.00 | -7.00 | |

BACKSCATTERING CROSS SECTION IS 1.8000 DECIBELS FOR SEA STATE FIVE

VERTICAL VERTICAL POLARIZATION

FREQUENCY OF INCIDENT WAVE IS 16.0 GIGAHERTZ

ANGLE OF INCIDENCE IN DEGREES IS 24.0 DEGREES

Figure 1. Backscattering Coefficients for Sea State Five and Radar Looking Downwind.


```

*****
*
*   BACKSCATTERING COEFFICIENT FOR DIFFERENT SEA STATES AND LOOKING CROSSWIND. OR
*
*   RADAR LOOKS IN THE CROSSWIND DIRECTION WITH SEA STATE SEVEN
*
*   BACKSCATTERING CROSS SECTION SIGMA VERSUS THETA FOR SEA STATE SEVEN
*
*   FREQUENCY   ANGLE OF INCIDENCE THETA IN DEGREES
*
*   IN GHZ      0 DEG      5 DEG      10 DEG      15 DEG      20 DEG      25 DEG
*
*   5.00        22.00      19.00      16.00      14.00      12.00      10.00
*
*   10.00       19.00      16.00      13.00      11.00      9.00       7.00
*
*   15.00       15.00      12.00      9.00       7.00      5.00      3.00
*
*   20.00       12.00      9.00      6.00      4.00      2.00      0.00
*
*   25.00       8.00      5.00      2.00      0.00     -2.00     -3.00
*
*   30.00       5.00      2.00     -1.00     -3.00     -5.00     -7.00
*
*   BACKSCATTERING CROSS SECTION IS 2.5000 DECIBELS FOR SEA STATE SEVEN
*
*   VERTICAL VERTICAL POLARIZATION
*
*   FREQUENCY OF INCIDENT WAVE IS 18.0 GIGAHERTZ
*
*   ANGLE OF INCIDENCE IN DEGREES IS 22.0 DEGREES
*
*****

```

Figure 2. Backscattering Coefficients for Sea State Seven and Radar Looking Crosswind.

```

*****
*
*   BACKSCATTERING COEFFICIENT FOR VEGETATION SUCH AS AGRICULTURAL CROPS, TREE
*
*   OR GRASSLANDS, AND BUSHES, SHRUBS, OR FLOWER COVER
*
*   BACKSCATTERING CROSS SECTION SIGMA VERSUS THETA FOR WHEATFIELDS
*
*   FREQUENCY   ANGLE OF INCIDENCE THETA IN DEGREES
*
*   IN GHZ      0 DEG      5 DEG      10 DEG      15 DEG      20 DEG      25 DEG
*
*   5.00        18.00      16.00      14.00      13.00      12.00      11.00
*
*   10.00       16.00      14.00      12.00      11.00      10.00      9.00
*
*   15.00       14.00      12.00      10.00      9.00      8.00      7.00
*
*   20.00       12.00      10.00      8.00      7.00      6.00      5.00
*
*   25.00       10.00      8.00      6.00      5.00      4.00      3.00
*
*   30.00       8.00      6.00      4.00      3.00      2.00      1.00
*
*   BACKSCATTERING CROSS SECTION IS 7.3200 DECIBELS FOR WHEATFIELDS
*
*   FREQUENCY OF INCIDENT WAVE IS 16.0 GIGAHERTZ
*
*   ANGLE OF INCIDENCE IN DEGREES IS 24.0 DEGREES
*
*****

```

Figure 3. Backscattering Coefficients for Vegetation in the Form of Wheatfields.

OBJECT ORIENTED DESIGN
DEMONSTRATED BY
SAR RAW DATA SIMULATOR DEVELOPMENT

A. Popella, R. Konjack, F. Leßke

German Aerospace Research Establishment (DFVLR)
WT-DA/MV SAR
Muenchnerstrasse 20,
D-8031 Oberpfaffenhofen, West Germany

Telephone (0049) 8153/28-1196
Telex 5270286 DAAL D
Telefax (0049) 8153/281137

Abstract

The German Aerospace Research Establishment (DFVLR) is currently proceeding in the development of the knowledge-based Intelligent SAR processor, ISAR, which will be used for high throughput and high precision routine processing of synthetic aperture radar image data from ERS-1 and X-SAR, being launched in 1990 and 1991 respectively.

The ISAR development is a challenging task, covering all aspects of a complex software package, which is distributed over several hardware systems and thus, is tackled using modern software-engineering methods.

The software design method, which has been chosen for the ISAR development, is the object oriented development. This design method has been applied to the development of a software based SAR raw data simulator, which is used first of all with respect to performance and engineering tests on the ISAR processor. Secondly, the simulator is used as a test instrument for applying modern software engineering standards, aiming at the elaboration of a method for accessing object oriented design, as well as tracing out the advantages in applying such a method.

The paper presents the project experiences in developing software, starting with system requirements, then functional decomposition and finally derivation of objects. This results in a modular software structure to be implemented in Ada. Particular consideration is given to the man-machine-interface.

Keywords: Object oriented software design, Ada, man-machine-interface,
SAR raw data simulation, ERS-1, X-SAR

NOISE MODELING AND ESTIMATION OF REMOTELY-SENSED IMAGES

J. S. Lee and K. Hoppel

Naval Research Laboratory, Washington, D.C. 20375-5000

ABSTRACT

A noise model is defined for remotely sensed images, and the noise statistics are estimated by using the means and variances from small (4x4 or 8x8) image blocks. Since most images contain many small but homogeneous areas, a scatter plot of variance vs (mean)² reveals characteristics of the noise. The Hough transform is then applied to the scatter plot to detect a straight line through the major cluster of data points. This defines the image noise statistics. Images from SAR, Landsat TM, and passive microwave sensors are used for illustration.

Keywords: noise modeling, SAR speckles, speckle statistics, noise filtering.

I. INTRODUCTION

Signal and noise modeling and filtering have always been important topics in digital signal processing. The signal is normally modeled as a random process which can be expressed as linear differential and difference equations driven by white noise [1]. The signal process can generally be identified by using auto-correlation techniques, time series analysis, or Fourier transform techniques [1]. The signal modeling of a digital image, however, is difficult due to the spatially-variant nature of the image. In this paper, we focus on modelling only the image noise without attempting to model the signal.

Noise appearing in remotely-sensed images makes image segmentation and computer scene description difficult. Noise filtering is commonly performed before further processing. Filtering noise effectively while retaining image quality requires a precisely-defined noise model. Early papers on noise filtering assumed without justification that the noise is additive [2]. Recently, a multiplicative noise model was developed for SAR images [3]. The multiplicative noise variance in SAR images depends on the number of looks used in the SAR processing. For both additive and multiplicative noise, the most frequently used method for estimating the noise parameters is to calculate statistics from homogeneous and featureless image areas located visually by the operator. In images containing rich detail, the absence of large homogeneous areas complicates the processes. A procedure is therefore presented which automatically models the image noise, and computes the noise parameters and distributions.

Based on the observation that most remotely sensed images contain many small homogeneous areas distributed over the entire image, the following algorithm is developed:

(1) The entire image is divided into small 8x8 or 4x4 blocks. The mean and variance are computed for each block.

(2) A scatter plot of the variance versus (mean)² reveals the characteristics of the noise. If the noise is additive, a large number of samples will cluster along a line of constant variance. If the noise is multiplicative, they will cluster along a sloped line through the origin. Some images may contain both additive and multiplicative noise. The latter can be gray level dependent. Blocks that are not homogeneous but which contain features (edges) will be sparsely scattered in the scatter plot with a higher value of variance.

(3) If two or more clusters are present in the scatter plot, they could be separated by gray level thresholding from the histogram.

(4) To detect a straight line through the center axis of a cluster while ignoring outlying edge blocks, the Hough transformation [2] is applied to each cluster.

(5) Noise distributions are obtained with pixels from the blocks in the major cluster.

Images from Landsat TM (Thematic Mapper), SIR-B (Shuttle Imaging Radar B), and passive microwave radiometers are used for illustration.

II. LINEAR NOISE MODEL

The motivation for constructing a noise model is to facilitate the noise filtering or to restore distorted images due to defocusing or motion blurring, etc.. Commonly assumed linear additive noise is described by:

$$z = x + w \quad (1)$$

where z is the image pixel gray level, x is the noise-free pixel level, and w is the additive noise with zero mean and standard deviation (S.D.), σ_w . Studies of SAR image speckle noise [3] have revealed a multiplicative noise form described by:

$$z = x v \quad (2)$$

where v is the multiplicative noise with unity mean and S.D. σ_v . The σ_v depends on the number of looks

used in the SAR processing and on whether the gray level represents the signal amplitude or intensity. For a 1-look amplitude SAR image, $\sigma_v = 0.5227$ and for a 4-look amplitude image, $\sigma_v = 0.26$. The algorithm developed here can be used to verify the type of processing used for a SAR image. When image enhancements have been performed, the noise characteristics are usually altered.

The general model proposed in this paper allows for both additive and multiplicative noise, described by:

$$z = xv + w \quad (3)$$

where x , v , and w are statistically independent. From a small homogeneous block the mean \bar{z} and variance $\text{var}(z)$ can be calculated by:

$$\bar{z} = \bar{x} \quad (4)$$

$$\text{and } \text{var}(z) = \sigma_v^2 \bar{z}^2 + \sigma_w^2 \quad (5)$$

The linear relation between $\text{var}(z)$ and \bar{z}^2 suggest a straight line fit in the scatter plot with slope σ_v^2 and intercept σ_w^2 . For the case of purely additive or purely multiplicative noise, it is easier and more precise to use a scatter plot of the standard deviation versus mean.

III. ESTIMATION ALGORITHM FOR NOISE STATISTICS

A commonly used method for estimating σ_v^2 and σ_w^2 is to identify large homogeneous image areas and use them to calculate the noise statistics. The problem with this method is that it must be supervised, the image might not have large homogeneous areas, and a large number of areas with varying means are necessary to define a linear noise model. A more appropriate unsupervised algorithm is to calculate $\text{var}(z)$ and \bar{z} from small blocks throughout the entire image. Assuming a large number of the small blocks correspond to homogeneous areas, a scatter plot of variance versus (mean)² should reveal a primary cluster. Blocks that contain features and edges will have a higher variance and be sparsely scattered above the primary cluster in the scatter plot.

In order to fit a straight line to the main cluster while excluding the outliers, the Hough transform is applied to the scatter plot. The maximum point in the Hough transform corresponds to the line passing through the maximum number of samples. A 1-look amplitude SAR image of Albemarle Sound N.C. (1024x1024 pixels), processed using 8x8 blocks is shown in Fig. 1(A). The corresponding scatter plot and straight line fit is shown in Fig. 1(B). The brightness of the pixels in the scatter plot is proportional to the number of samples at that position. The additive noise variance from the straight line model is very small. The multiplicative noise variance σ_v^2 is 0.249 ($\sigma_v = 0.5$) which is very close to the theoretical value of 0.522 [3], indicating the multiplicative nature of the 1-look SAR image. The same scene using 4-look amplitude processing is shown in Fig. 2(A) and the corresponding scatter plot in Fig. 2(B). The straight line model gives a $\sigma_v^2 = 0$ and $\sigma_v^2 = 0.0875$ ($\sigma_v = 0.29$) only slightly larger than the theoretical σ_v of 0.26.

Errors in the estimated σ_v could be contributed to by the following factors. (1) The small block size creates a sampling error; (2) In order to apply a

reasonable size Hough transform the scatter plot (256² by 256²) is quantized into cells. The large quantization error would effect the accuracy of the Hough transform. This error is reduced if the scatter plot of standard deviation versus mean is used (256 by 256), but combined additive and multiplicative noise would not cluster along a line. (3) The Hough transform is an easy way to achieve the straight line fit while excluding outlying samples. There may be more accurate ways such as using a least squares fit to the samples within a fixed neighborhood of the line identified by the Hough transform.



(A) One-look SIR-B IMAGE
(Albemarle Sound, NC.)



(B) Scatter plot to define the noise (multiplicative noise are detected)

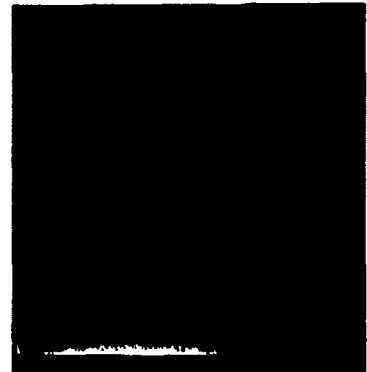
Fig. 1 Speckle noise is defined as multiplicative noise with S.D.=0.5 as compared with the theoretical value of 0.5227. The 8x8 block size is used. Each unit in (B) is (64)².

IV. OTHER EXAMPLES

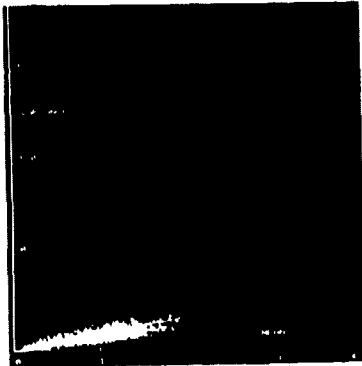
A Landsat TM image (band 4) of the Washington D.C. area is shown in Fig. 3(A). The noise level is low ($\sigma_w = 8$) and has the characteristics of additive noise. To demonstrate the effectiveness of the algorithm, gaussian additive noise with $\sigma_w = 30$ is added to the image, Fig 3(C&D). The linear model gives $\sigma_v \approx 0$ and $\sigma_w^2 = 831$ ($\sigma_w = 28.8$). Theoretically, the σ_v^2 should be 964. Possible reasons for this discrepancy were given in section III. For the purposes of noise filtering [3] and segmentation [4], the results obtained are sufficiently accurate.



(A) Four-look SIR-B image

(A) Landsat TM-Band 4
(Washington D.C.)

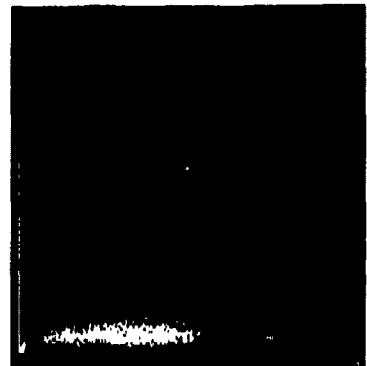
(B) Scatter plot



(B) Scatter plot of (A)



(C) Noise added to (A)



(D) Scatter plot shows additive noise

Fig. 2 Speckle noise is defined as multiplicative noise with S.D. = 0.29 as compared with the theoretical value of 0.26. The 8x8 block size is used. Each unit in (B) is $(64)^2$.

Fig. 3 Additive noise modeling is demonstrated by simulating an additive noise image (C) with noise S.D. = 30.

A passive microwave image and noise model are shown in Fig. 4(A). The noise can be observed to be scan line noise due to an instrumental defect. The scatter plot, Fig. 4(B), is of the standard deviation versus mean for better resolution. Two clusters indicate that two classes (land and water) exist in the image and could be segmented by a threshold value of 128. Although a noise model could be estimated for each class, a single line has been fitted. The noise is principally additive with $\sigma_w = 79.7$.

V NOISE DISTRIBUTION ESTIMATION

The noise probability distribution can also be estimated. This is done by choosing a small box on the line detected in the scatter plot (Fig. 5(A)), which defines a range of $(\text{mean})^2$ and variances. All of the image blocks which fall into that range are used to compute the histogram or distribution of image pixels. This method is illustrated in Fig. 5 using the SAR 1-look and 4-look images from Fig. 1 and 2. The histograms (dots) are shown along with the theoretical distributions [3] (solid curve). The theoretical distribution for the 1-look amplitude is the Rayleigh distribution and for the 4-look amplitude image it is a Chi distribution. The mean value used

for the theoretical distributions is the mean value of all the pixels used in the histogram. The reasonably good fits substantiate the effectiveness of the algorithm and verify the theoretical distributions.

VI. CONCLUSIONS AND DISCUSSION

The primary reason for obtaining noise characteristics from the scatter plot is for subsequent use for image filtering. Nonlinear image filtering and restoration is a very complicated process and yields questionable accuracy. On the other hand, the sigma filter [5] for image noise smoothing based on the present model, can be easily modified for a nonlinear noise model. Other algorithms for noise smoothing could be modified to operate in a piecewise linear fashion.

The noise modeling and noise statistics estimation algorithm developed in this paper work reasonably well for images corrupted by both additive and multiplicative noise. For SAR images in particular, this algorithm can be used to automatically distinguish between single and multilook processing.

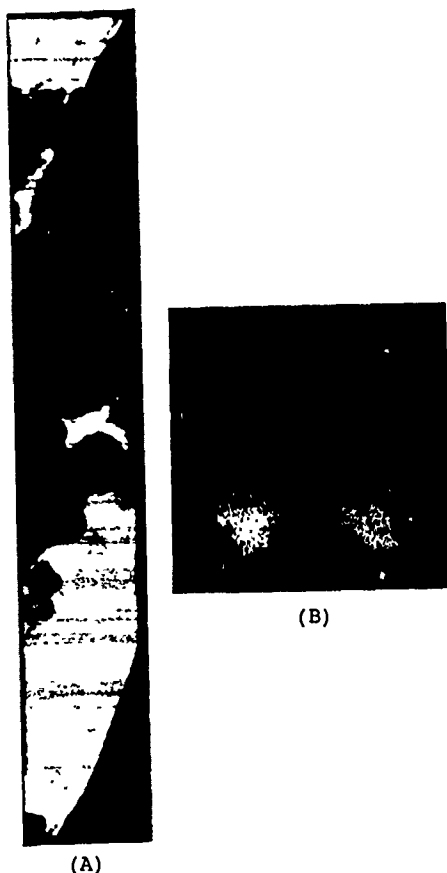
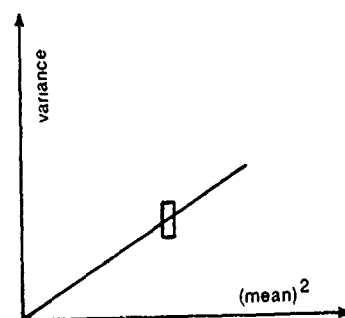


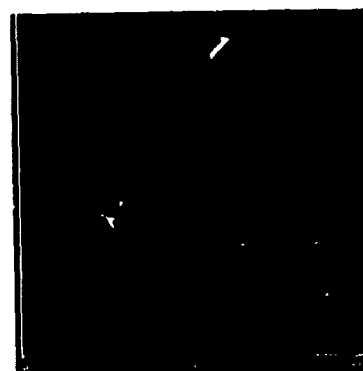
Fig. 4 Passive microwave image is shown in (A). Two clusters in the scatter plot as shown in (B) indicates the separation of land and water pixels. The additive noise is detected with $S. D. = 29.7$.

REFERENCES

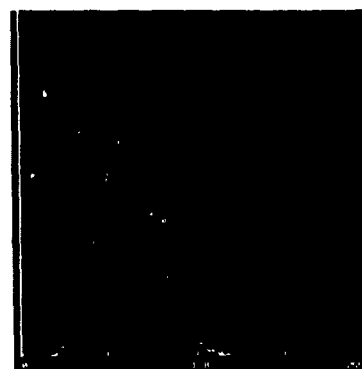
- [1] Gelb, A. Applied Optimal Estimation, Chapter 3, Mass Inst. of Technology Press, 1974.
- [2] Pratt, W. K., Digital Image Processing, Wiley, New York, 1978.
- [3] J. S. Lee, "Speckle Suppression and Analysis for Synthetic Aperture Radar Images", Optical Engineering, 636-643, 1986.
- [4] Lee, J. S. & Jurkevich, I., "Segmentation of SAR Images", Proceedings of IGARSS '88, p. 1503-1507, 1988.
- [5] Lee, J. S. "Digital Image Smoothing and the Sigma Filter", Computer Vision, Graphics and Image Processing 23, 255-269, 1983.



(A) Define a small neighborhood as shown in box



(B) Histogram and p.d.f. of 1-look SAR



(C) Histogram and p.d.f. of 4-look SAR

Fig. 5 Noise probability density function (p.d.f.) estimation of SAR amplitude images. The normalized histogram is shown in dots, while the Rayleigh distribution for 1-look SAR and Chi distribution for 4-look SAR are in solid lines.

Simulation of Sub-Pixel Terrain Effects on Radar Backscattering of Snow

Jianchen Shi and Jeff Dozier

University of California, Center for Remote Sensing and Environmental Optics
Santa Barbara, CA 93106

ABSTRACT

The purpose of this study is to examine the dependence of the radar backscattering coefficient σ^0 on terrain geometry for a rough surface at subpixel resolution. Of particular interest are the variations in σ^0 caused by different angles of incidence within the pixel and the determination of the local angle of incidence for a pixel when the topographic features change within it. The effect of terrain geometry on the radar backscattering coefficient at the subpixel scale is evaluated by the facet model under incoherent summation. Based on the backscattering coefficients simulated from the facet model, the averaged equivalent-angle method provides a useful way to determine the local angle of incidence for the pixel.

INTRODUCTION

In radar images of alpine snow, we are interested in the relationships between backscattering coefficients and snowpack parameters, such as density, grain size, depth, and liquid water content. The backscattering coefficient σ^0 also depends on the local angle of incidence, which is usually assumed to be the same for all points within a resolution cell, i.e. a pixel. In mountainous areas, however, topographic features often change within a radar image pixel. Since a rough terrain can be approximated through a series of small planar facets, each tangential to the actual surface, an illuminated pixel may be evaluated as the combination of many subpixels, each having different gradients. The availability of a high resolution Digital Elevation Model (DEM) makes it possible to simulate the effect of terrain geometry on the radar backscattering coefficient for a random rough surface at subpixel resolution (Figure 1).

The test samples were selected from a DEM with 5-m resolution and better than 1 m vertical accuracy for Emerald Lake in Sequoia National Park, California [1]. The geometric parameters within the 25-m pixels cover a range of the standard deviation of the slope angles up to 7° and the variation of the aspect angles which is measured in the mean resultant length is up to 0.93 in a pixel cell. The mean resultant length is a measurement of dispersion analogous to the variance, and ranges from zero to one. That is, large values indicate that the observations are tightly bunched together with small dispersion [2]. The viewing angle is from 0° to 55° for every 5° in nadir and from 0° to 355° for every 5° in azimuth. To simulate the terrain effect on radar backscattering at subpixel resolution, we used experimental data [3] for the backscattering coefficients for wet and dry snow at different incidence angles, 8.6 GHz frequency and HH polarization.

FACET MODEL

The radar backscattering coefficient σ^0 of a resolution cell is equal to the radar cross section of the cell normalized by its cross-sectional area. Since σ^0 represents an averaged value, it may be determined by incoherent summation of the facet backscattering $\sigma_{A_i}^0$ from each subpixel ΔA_i in the resolution cell.

$$\sigma^0[\theta, \phi] = \frac{\sum_{i=1}^n \sigma_{A_i}^0[\theta_i(\theta, \phi)] \times \Delta A_i}{A} \quad (1)$$

ΔA_i is an illuminated subpixel surface area, and A is the area of the pixel. $\sigma_{A_i}^0[\theta_i(\theta, \phi)]$ is the backscattering coefficient at local angle of incidence $\theta_i(\theta, \phi)$ of each subpixel ΔA_i . $\theta_i(\theta, \phi)$ is a function of the radar azimuth angle ϕ , nadir angle θ , slope angle S_i , and aspect angle E_i of each subpixel ΔA_i . It can be determined by:

$$\cos \theta_i(\theta, \phi) = \cos \theta \cos S_i + \sin \theta \sin S_i \cos(\phi - E_i) \quad [2]$$

This represents the backscattering coefficient σ^0 of a unit area as an averaged value of the combination or summation from the backscattering coefficients of each subpixel within it. The accuracy of σ^0 obtained from the Facet Model depends on the resolution of the DEM data. The higher the DEM resolution is, the better approximation can be made for σ^0 .

SIMULATION OF TERRAIN EFFECTS

In order to assess the effect of subpixel topography on the backscattering coefficient, three configurations will be evaluated:

1. The gradient (S, E) of a 25-m pixel is determined from a DEM with 25-m grid resolution.
2. The gradient (S, E) of a 25-m pixel is determined from mean slope and aspect angles from 5-m DEM (the reference plane method).
3. The surface geometric parameters for each subpixel are determined from a 5-m DEM, and the facet model is used to calculate σ^0 of a 25-m pixel.

Figure 2 shows the backscattering coefficients of a 25-m pixel simulated under the above three configurations by using the measured data for wet snow [3]. Figure 3 shows the absolute difference in backscattering coefficients for both wet and dry snow from the reference plane method compared with the facet model. The backscattering coefficients from configurations 1 and 2 could be considered as what they should be at the angles of incidence under those configurations. Obviously, using the same resolution DEM data to determine the geometric parameters could result in misinterpretation when the geometric parameters change greatly at subpixel scale. On the other hand, using the averaged geometric parameters from the higher resolution DEM, this situation can be improved. But this method might still cause the significant error when there is near normal incidence, that is when radar nadir angle is approximately equal to the mean slope angle and radar looks nearly directly on the slope. The slopes of the angular response curves are generally much greater at near normal incidence than at large angles of incidence. Since the slopes of the angular response curves from wet and dry snow are also different, the errors are different for the different optical surfaces at the same level of misinterpretation of the local angle of incidence.

DETERMINATION OF THE EQUIVALENT ANGLE OF INCIDENCE FOR A PIXEL

Since rough topography affects the distribution of incidence angles for a pixel, which, in turn, affect the interpretation of radar backscattering coefficients, the simulation depends on determining an angle of incidence within a pixel for a rough terrain.

Figure 3. Effect of Terrain Geometry at Different Nadir Angles

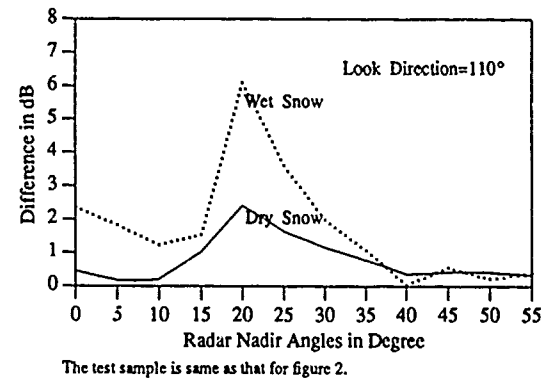


Figure 5. Compared with Facet Model at Different Nadirs

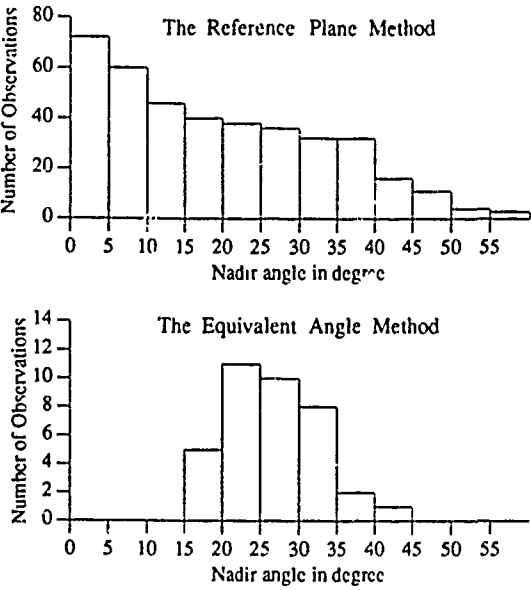


Figure 4. Compared Two methods with the Facet Model

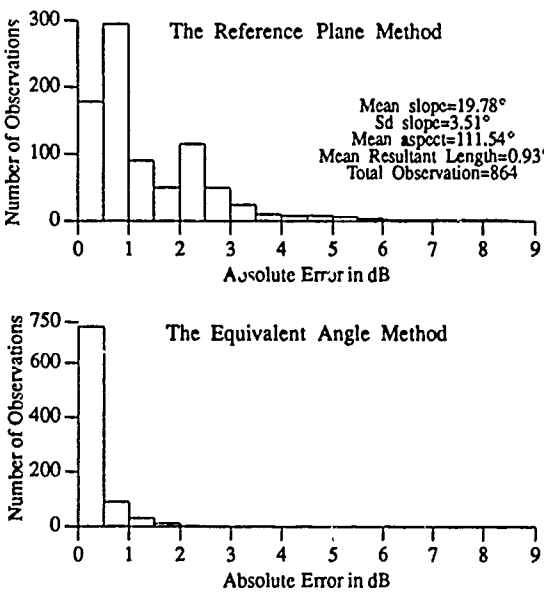


Figure 3. Effect of Terrain Geometry at Different Nadir Angles

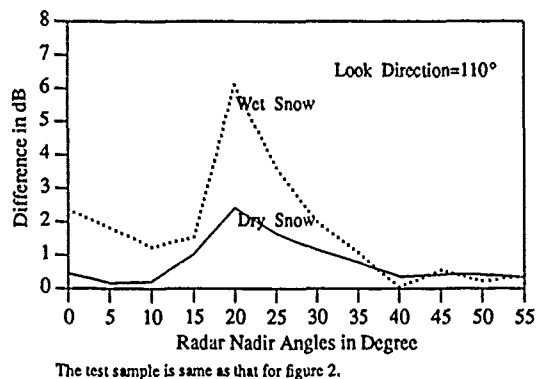


Figure 4. Compared Two methods with the Facet Model

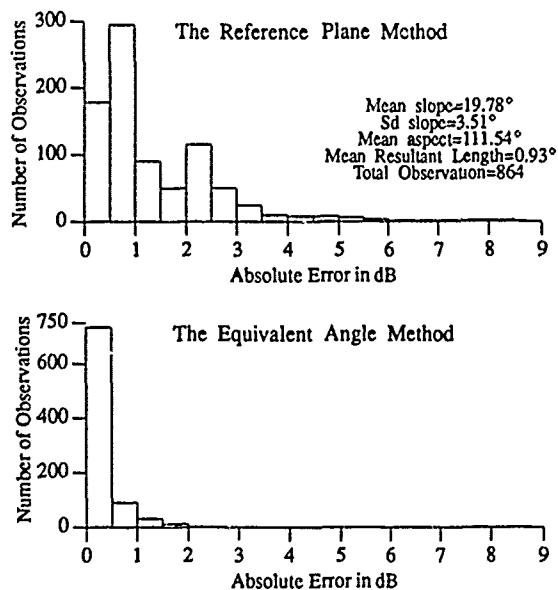
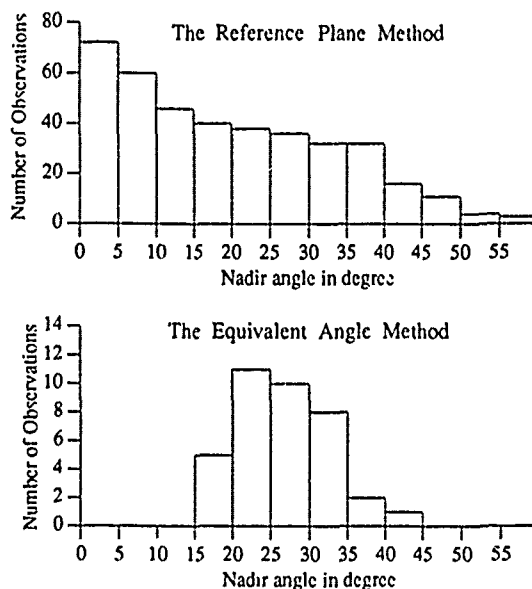


Figure 5. Compared with Facet Model at Different Nadirs



We define an equivalent angle of incidence θ^* for the pixel, at which the radar backscattering coefficient $\sigma^0(\theta^*)$ is the same as that from a flat surface $\sigma^0(\theta)$ with the same optical properties. Based on simulated backscattering coefficients from the facet model, the equivalent angle of incidence can be found by using the inverse method on the same angular response curve.

$$\sigma^0(\theta^*) = \sigma^0(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\sigma_{A_i}^0[\theta_i(\theta, \phi)]}{\cos \theta_i} \quad [3]$$

From the facet model, we can see that the shape of the angular response curve (backscattering coefficient vs angle of incidence) is important only to determine the equivalent angle at a given range of the local angles of incidence at each subpixel. Snowpack parameters which affect the backscattering coefficients change in alpine snow covered drainage basins also. We thus might use an averaged equivalent angle of incidence from the two angular response curves for wet and dry snow to determine the equivalent angle of incidence for the pixel without regard to the change in snowpack parameters. Figure 4 shows the absolute error terms compared with those obtained from the facet model by using two methods: (1) the reference plane method, (2) the averaged equivalent-angle method. It is clear that using the equivalent-angle method is much better than using the reference plane method to determine an equivalent angle of incidence for a pixel when its topography changes within a subpixel scale.

DISCUSSION

Table 1 shows the number of observations exceeding the critical point and the maximum absolute error in a total of 864 observations of each test sample when we use the equivalent-angle method from the two snow measurement data to correct the backscattering coefficients, compared with those obtained from the facet model, for wet snow under a variety of geometric parameters. The critical point for absolute error term was set to 1 dB in our study, which corresponds to the relative calibration accuracy for the Shuttle Imaging Radar - C (SIR-C) [4].

| Std of slope | Mean Resultant Length | | | | | |
|--------------|-----------------------|------|------|------|------|------|
| | 0.97 | | 0.95 | | 0.93 | |
| | No. | Max. | No. | Max. | No. | Max. |
| 3.0° | 0 | 0.58 | 0 | 0.64 | 21 | 1.24 |
| 3.6° | 0 | 0.74 | 22 | 1.54 | 21 | 1.55 |
| 4.0° | 5 | 1.26 | 30 | 1.68 | 69 | 3.54 |
| 4.5° | 24 | 2.67 | 71 | 2.26 | 55 | 1.75 |
| 5.7° | 15 | 1.32 | 63 | 3.83 | 92 | 2.43 |
| 7.0° | 80 | 1.64 | 73 | 2.97 | 178 | 1.84 |

TABLE 1. Absolute error by using equivalent-angle method

Generally, the accuracy of the equivalent-angle method decreases when the variation of the surface geometric parameters in the pixel increases. But the relationship is also affected by the distribution of the surface geometry. For the same standard deviation of slope angle and mean resultant length, the error is greater from the bimodal distribution, such as a test sample straddling a ridge, than from a Gaussian or uniform distribution. Figure 5 shows the number of observations that exceed the critical point, by both the equivalent-angle method and the reference plane method for wet snow. By using the reference plane method, the error terms mainly occur at nadir and decrease when nadir angle increase. On the other hand, by using the equivalent-angle method, the error occurs mainly under the situation that radar nadir angle is approximately equal to the mean slope angle and radar looks nearly direct on the slope. However, the error is much less by the equivalent-angle method than by the reference-plane method.

CONCLUSIONS

The results show that the terrain effect on radar backscattering at subpixel scale is determined by two sets of parameters: (1) geometric parameters, which include the standard deviation of the slope and aspect angles within a pixel cell, and look geometry, which includes the radar nadir angle and azimuth view; (2) surface optical properties, which determine the shape of the angular response curve.

Based on the backscattering coefficients simulated from the facet model, the averaged equivalent-angle method provides better accuracy, than the reference plane method. The validity can be determined if the difference, in the averaged equivalent angle and any of the equivalent angles obtained from the two extreme angular response curves, results in the backscattering coefficients exceed the critical point. The validity also depends on the

shape of the two extreme curves, and it is a function of the angle of incidence for the given angular response curves.

If we use two extreme angular response curves with the most difference in shape from two field measurements for a study area, and the validity is determined from the two extreme angular response curves, the actual error terms should be smaller than those from predicting two extreme cases since other types of snow in the study area falls between the two extreme cases and the equivalent angle, again, must be between those obtained from the two extreme curves. In other words, the equivalent angle from other types of snow in the study area is closer to the averaged equivalent angle than that from the two extreme curves.

REFERENCES

- [1] J. Dozier, J. M. Melack, D. Marks, K. Elder, R. Kattelmann, and M. Williams, "Snow, snow melt, rain, runoff, and chemistry in a Sierra Nevada watershed," Final report, Contr. A6-147-32, Calif. Air Resour. Board, 1989.
- [2] J. C. Davis, *Statistics and Data Analysis in Geology*, pp. 314-321, Kansas Geological Survey, 1986.
- [3] F. T. Ulaby, W. H. Sules, and M. Abdelrazik, "Snowcover influence on backscattering from terrain," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-22, no. 2, pp. 126-132, 1984.
- [4] National Aeronautics and Space Administration, *Shuttle Imaging Radar-C Science Plan*, JPL Publ. 86-29, Pasadena, CA: Jet Propulsion Laboratory, 1986.

Figure 1. Geometry Configuration - Terrain through Facet Representation

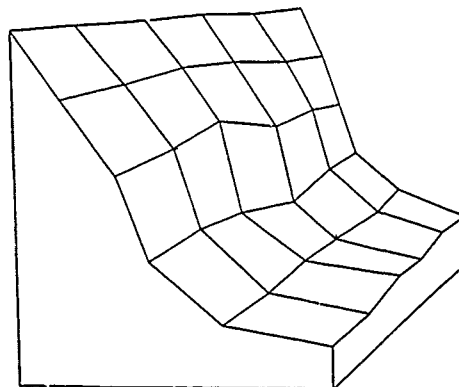
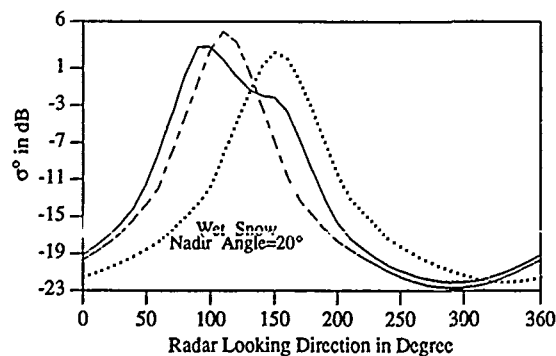


Figure 2. Effect of Terrain Geometry at Different Look Directions



The dotted, dashed and solid lines represent configuration 1, 2, and 3, respectively. The geometry parameters within the test sample are: mean slope angle = 19.78°, standard deviation of slope angle = 3.51°, mean aspect = 111.54°, and the mean resultant length = 0.93.

CALCULATION OF THE SPATIAL DISTRIBUTION OF SCATTERERS IN A DIFFUSE SCENE FROM SAR DATA

Christopher C. Wackerman

Radar Science Laboratory
Advanced Concept Division
Environmental Research Institute of Michigan
Ann Arbor, MI 48107 USA

1.0 Introduction

One practical use of synthetic aperture radar (SAR) images is to segment large areas of land into various geophysical classes; crop types, tree types, or urban versus suburban areas for example. Since large amounts of data are usually involved automatic algorithms are the only practical approach. It would make such algorithms easier to implement if the probability density functions (pdf's) of the various classes could be modeled beforehand; in fact if the pdf's are known optimal detection algorithms can be implemented. In addition, knowledge about the distribution of scattered energy from a given scene can give information about the physical structure of the scattering surface, namely the spatial distribution of scatterers within the scene. For reasons such as these it is useful to develop a parametric model for the pdf of diffuse scenes in SAR images.

Much work has been done in this area already [1,2,3], mainly in attempting to fit pdf's with known analytical forms to histograms generated from SAR data. Unfortunately, as SAR resolutions become better (thus generating more independent samples of a diffuse scene) and SAR calibration procedures become more refined (thus removing uncertainties in backscattered values due to system effects) the analytical forms do a poorer job of fitting the actual histogram data. This indicates that a more complicated pdf model is necessary, one that includes the physical aspects of the terrain being imaged, rather than just continuing to search for other analytical forms. This paper will present a model that essentially divides the SAR pdf into a part due to speckle and a part due to the spatial distribution of the scatterers within the scene allowing a better fit to the SAR histogram values and allowing the spatial distribution function to be extracted from the SAR data. A specific form of this model was presented in ref. 2, but our approach is to generalize that model to multi-looked SAR data and to extract the spatial density function directly in addition to assuming a functional form for it.

2.0 Density Function Model

To form a simple scattering model for diffuse targets in SAR imagery, we will divide the target spatially into small cells on the order of the resolution of the SAR. We will assume that each cell contains a large number of scatterers at ranges that vary much more than the wavelength of the SAR and that all have the same radar crosssection. However, the

radar crosssection values will vary from cell to cell. Because the surface within each cell will appear rough to the radar, the backscattered values will fluctuate due to speckle with an exponential density function [5] (if we are considering intensity statistics). If the SAR data has been multi-looked, then the statistics of the speckle will be the Gamma distributed with density function $\Gamma(x;c,b)$;

$$\Gamma(x;c,b) = (x/c)^{c-1} [\exp(-x/b)] / b \Gamma(c) \quad (1)$$

where $\Gamma(c)$ is the Gamma function, c will be an integer representing the number of independent speckle values that were added together and $b=(\sigma/c)$ where σ is the mean radar crosssection for that cell. Let $f_{\sigma}(\sigma)$ be the density function that represents how the mean radar crosssections vary from cell to cell; i.e. the spatial distribution of the scatterers. Considering the SAR backscattered values from the entire target as a single random variable, it will then have density function $f(x)$;

$$f(x) = \int_0^{\infty} \Gamma(x;c,\sigma/c) f_{\sigma}(\sigma) d\sigma \quad (2)$$

In Eq. (2), the Gamma density represents the variation due to speckle and the f_{σ} density represents the variation due to spatial changes in radar backscatter.

The actual measurement on the data is a histogram of the SAR values within the target. Let h_i represent the histogram values for $i = 1, \dots, B$ where B is the number of bins in the histogram. Then

$$h_i = \frac{\text{number of pixels } \in (S_i, E_i)}{\text{total number of pixels in the target}} \quad (3)$$

and

$$h_i \approx \int_{S_i}^{E_i} f(x) dx \quad (4)$$

where (S_i, E_i) , $i = 1, \dots, B$ represents the bin structure with which the histogram was calculated. Our problem can now be stated as follows: given values of h_i , $i = 1, \dots, B$ determine $f_{\sigma}(\sigma)$.

If we consider h_i as an estimate of $f(x_i)$ for some x value in the middle of the bin (S_i, E_i) , then our problem is solving the first order integral equation (2). Unfortunately, such equations are known to be ill-posed and when put into the form of a system of linear equations (which we would need to do since

we only have sampled data) generate ill-conditioned matrices which cannot be inverted. Our solution to this difficulty is two-fold. First, assume an analytical form for f_σ and then solve for the parameters. This has the added advantage of allowing us to use Eq. (4) directly instead of the approximation mentioned above. This is similar to an analysis in [2], but we use more general density functions and find an optimal solution in the sense that it minimizes an error metric. Second, we will perturb the analytical solution to f_σ iteratively to minimize an error metric. This allows us to find a solution that is not constrained to any special form, but gets around the ill-conditioning mentioned above since we are finding a solution that is close to an initial guess.

3.0 Analytical Form Fitting

We will assume that f_σ can also be described as a Gamma density function; i.e. $f_\sigma(\sigma) = \Gamma(\sigma; c_0, b_0)$ where c_0 and b_0 need to be determined. This choice is prompted by the wide range of density shapes that the Gamma can have for different values of c_0 and b_0 ; from exponential to Rayleigh-like to Gaussian. Also, it will allow us to determine the general shape for f_σ ; we can refine that shape with the perturbation method discussed below.

Substituting our Gamma assumption into Eq. (2) generates a three parameter model for $f(x)$ with the three parameters being c , c_0 and b_0 . In theory, we should know the value of c a priori since it corresponds to the number of looks used to generate the SAR image. In practice however, due to the different methods that are used to generate multi-looked data and the possibilities of different weights being applied to each individual look, we usually can only estimate c from the methodology used. Thus we have decided to leave it as a parameter that needs to be specified. The moments about the origin for this model can be easily calculated;

$$E[x^r] = \left(\frac{b_0}{c}\right)^r \sum_{k=0}^{r-1} \frac{\pi}{c} (c+k) (c_0+k) \quad (5)$$

and specifically the mean = $b_0 c_0$ and the variance = $b_0^2 c_0 (1 + [c_0 + 1]/c)$.

The simplest method to implement for finding the values of the three parameters that best fit a SAR image is trial and error. Although not optimal in any sense, it does allow a quick analysis for the general shape that f_σ has and allows generation of an initial guess for the iterative procedures discussed below. To that end we analyzed how changes in c , c_0 and b_0 actual perturb $f(x)$. We found it easier to work with the distribution function for this analysis than the density function $f(x)$. Figure 1 shows how the distribution function changes with the c parameter; note that it is essentially a rotation of the distribution function. Figure 2 shows how changes in c_0 perturb the distribution function; it essentially shifts the function. The b_0 parameter also shifted the distribution function similar to the c_0 parameter. Using mean = $b_0 c_0$ allowed us to eyeball fits to various SAR images. Figure 3 shows an actual SAR image distribution function (circles) compared to an eyeball fit to our model (solid line). For comparison a fit to a simple Gamma model (using the methods of

moments to estimate the parameters) is also shown (dotted line). Note that although both models do a good job the single Gamma model slightly under predicts the distribution values at the first turn and then over predicts at the second turn whereas the Gamma-Gamma model does a good job of predicting at the first turn and only slightly over predicts at the second turn. The mean squared error of the Gamma-Gamma model in Figure 3 was 30% lower than for the single Gamma model.

For a more optimal fit, although at some computational cost, we iteratively found the parameters c , c_0 , b_0 that minimized the mean square error metric E ,

$$E = \frac{1}{B} \sum_{i=1}^B (h_i - f_i)^2 \quad (6)$$

using a Levenberg-Marquardt method [5]. To determine any improvement this model gave over fitting to a single analytical form, the algorithm was run on a series of SAR image subsets which represented different ice types. A single Gamma density function was then fit to the same data using the same algorithm; although this time only finding the value of two parameters that minimized E . Due to computational costs only a few such cases could be run, but in each instance the value of E for the Gamma-Gamma model was 30 to 50 percent better than the single Gamma model. To avoid the problem of comparing a three parameter model (i.e. Eq. (2)) with a two parameter model (the single Gamma fit) we also ran the algorithm by setting c to an estimate of the number of looks of the images and allowed only c_0 and b_0 to vary. The were essentially similar to the previous case with the improvement in mean square error being approximately the same.

4.0 Perturbation of the Solution

The analysis discussed above assumes a shape for the f_σ density function. To avoid this we analyzed perturbing the Gamma solution iteratively to determine if a different shape could generate a smaller value of E . Theoretically we could apply the same iterative algorithm as above, however the dimensionality of the resultant parameter space makes the method not practical to implement. Instead, we perturbed each f_i separately and determined which value minimized E . Performing this for each value of i separately, we then put them together to generate our "next guess" for f_σ . Although this is not an optimal guess, and for special cases can cause a stagnation, for most well behaved functions it will move the estimate closer to the value that minimizes E and it will allow an estimate of the actual shape for f_σ . Figure 4 shows an original guess for the f_σ Gamma density function as generated in section 3 (solid line) and the result of performing one iteration of the perturbation analysis (connected circles). Note that the algorithm is attempting to push the density function into a shape that is more delta-like with much heavier tails than the Gamma shape; perhaps indicative of less variation in the spatial scatterers than the Gamma curve indicated. Implementation of this algorithm on an array processor is currently being performed to carry this analysis further.

5.0 Conclusions

We have presented a model for the density function of SAR images of diffuse scenes that is a first order attempt at modeling the physical scattering properties and does a better job of fitting to actual SAR image histograms than fitting single analytical forms. In addition, the resultant parameters allow better segmentation of different SAR scattering phenomena since they remove the variation within classes caused by speckle. Initial results indicate that a perturbation of the Gamma solution to Eq. (2) appears to be possible and generates a more delta-like shape to the density function.

6.0 Acknowledgments

This research was supported by funds from the Office of Naval Research (ONR), under Contract No. 0014-81-C-0295. The ONR technical monitor was Charles A. Luther. The author would like to thank Rita Woods and Janice Anquetil for their help in preparing the manuscript.

7.0 References

- [1] Lyden, J., B. Burns, A. Maffett, "Characterization of sea ice types using synthetic aperture radar," IEEE Trans. Geoscience and Remote Sensing, vol. GE-22, No. 5, pp. 431-439, Sept. 1984
- [2] Zito, R., "The shape of SAR histograms," Compt. Vision Graph. Image Proc., vol. 43, pp 281-293 1988
- [3] Kasischke, E., A. Maffett, R. Larson, "Statistical modeling of intensity distribution on airborne SAR imagery," Proc. of IGARSS '87, pp. 1357-1362, May 1987
- [4] Goodman, J., "Some fundamental properties of speckle," J. Opt. Soc. Am. vol. 66, No. 11, pp.1145-1150, Nov. 1976
- [5] Press, W. et. al., Numerical Recipes, Cambridge University Press, 1986, pp. 523-526

Distribution Function
Changes with C Parameter

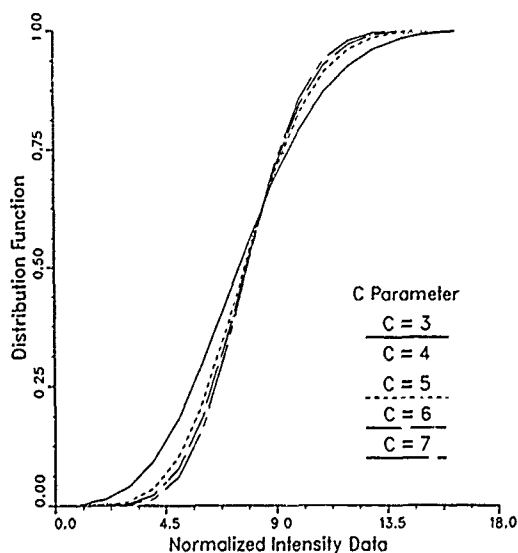


Figure 1

Distribution Function
Changes with c_0 Parameter

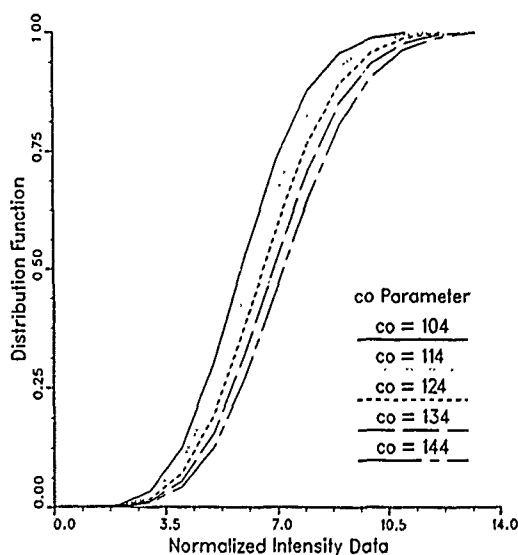


Figure 2

Comparison of Model to
Standard Gamma

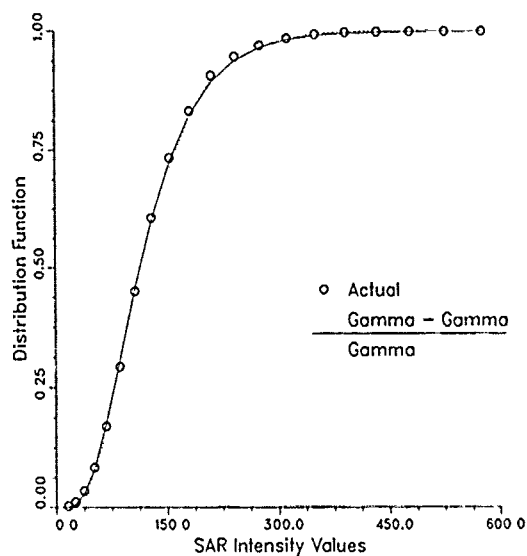


Figure 3

Comparison of Gamma-Gamma
Solution to Perturbed Solution

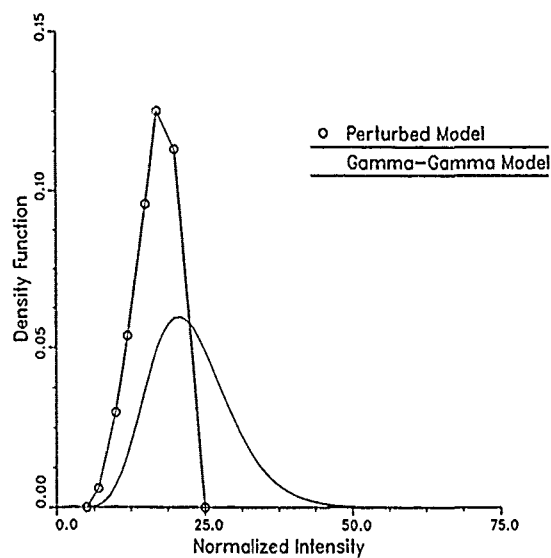


Figure 4

COMBATTING SPECKLE IN SAR IMAGES: VECTOR FILTERING AND SEQUENTIAL CLASSIFICATION BASED ON A MULTIPLICATIVE NOISE MODEL *

Qian Lin and Jan Allebach

Laboratory for Applications of Remote Sensing and
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

ABSTRACT

We propose an adaptive vector linear minimum mean-squared error (LMMSE) filter for multichannel images with multiplicative noise. We show theoretically that the mean-squared error in the filter output is reduced by making use of the correlation between image bands. The vector and conventional scalar LMMSE filters are applied to a 3 band SIR-B SAR image, and their performance is compared. We also show that the speckle noise distribution of an N-look amplitude SAR image is square-root-of-gamma. Based on this distribution and a multiplicative noise model, we derive the per-pel maximum likelihood classifier. We extend this to the design of sequential and robust classifiers. These classifiers are also applied to the 3 band SIR-B SAR image.

Keywords: synthetic aperture radar, speckle, vector filter, contextual classifier.

1. INTRODUCTION

Because of its ability to operate under all weather conditions, synthetic aperture radar (SAR) is playing an increasingly important role in gathering information from the earth's surface. At the same time, new instrumentation with expanded capabilities is being developed. It is becoming common for SAR images to have multiple bands corresponding to different incidence angles, different polarizations, and different frequencies, or a combination of these. For example, in the Shuttle Imaging Radar-C experiment scheduled for 1991, SAR images will be acquired simultaneously at two microwave frequencies (L and C band) with four polarizations (HH, VV, HV, and VH). However, SAR is a coherent system; and images acquired with it are corrupted by speckle noise. This impedes visual interpretation, and severely degrades the performance of per-pel classifiers. To reduce speckle noise, multilook processing, in which independent frames are averaged, is applied to the SAR images [1]. The resulting noise has been shown to fit a multiplicative model [2].

In previous work in this area [2]-[5], numerous filters have been designed for further speckle reduction. These are based on various models for speckle, including multiplicative noise and non-Gaussian statistics. However, the image bands are filtered separately and independently. The correlation between bands is not utilized. Recently, a multichannel filter was proposed for stationary images corrupted by additive noise [6]. In classifying ground cover type, the per-pel maximum likelihood (ML) classifier, based on an additive Gaussian noise model, cannot perform satisfactorily due to speckle

noise. One approach to improving classifier performance is to filter the image first, and then apply the per-pel classifier. Another approach is to use contextual information in designing the classifier. In our previous research and that of others [7] [8], it has been shown that both approaches yield a significant improvement in classification rate. Among the filters we investigated, the scalar adaptive linear minimum mean-squared error (LMMSE) filter [2] was one of the best in preserving edges and roads, and in improving classifier performance.

In this paper, we exploit the correlation between image bands to aid in speckle reduction. We propose an adaptive vector filter for nonstationary images, which is a generalization of the adaptive scalar LMMSE filter. We assume that the noise is multiplicative, and uncorrelated from band to band. We derive the LMMSE estimate of the vector of terrain reflectances in terms of the SAR image and noise statistics. SIR-B images of a forested area in Florida are filtered with the new vector filter, and compared with images in which each band is filtered separately with the scalar adaptive LMMSE filter.

The conventional per-pel ML classifier implicitly assumes an additive Gaussian noise model. In this paper, we propose a per-pel ML classifier based on a multiplicative noise model. We investigate the speckle noise distribution of an N-look SAR amplitude image, and derive the ML estimate of the class statistics, and the log-likelihood function based on this model and noise distribution.

We previously developed a contextual classifier based on sequential decision theory and an additive Gaussian noise model [7]. In this paper, we use a generalized sequential probability ratio test to design a classifier based on the likelihood function for the multiplicative noise model. To limit the effect of deviations from the assumed model on classifier performance, we also design a robust sequential classifier for multiplicative noise. We compare the performance of these classifiers using the SIR-B data.

2. ADAPTIVE LMMSE VECTOR FILTER

The adaptive vector LMMSE filter is based on a model in which the noise is multiplicative, unit mean, and uncorrelated. The gray value z_i in band i of the SAR image is related to the corresponding terrain reflectance x_i and noise v_i according to $z_i = x_i v_i$, $i = 1, 2, \dots, M$, where M is the total number of bands. In vector notation, let $\vec{z} = [z_1 \ z_2 \ \dots \ z_M]^T$, and $\vec{x} = [x_1 \ x_2 \ \dots \ x_M]^T$. The general form for the LMMSE estimate of the terrain reflectance \vec{x} is

$$\hat{\vec{x}} = \vec{\mu}_x + \pi_{xx} \pi_x^{-1} (\vec{z} - \vec{\mu}_z),$$

where $\vec{\mu}_x$ and $\vec{\mu}_z$ are the means of \vec{x} and \vec{z} , π_{xx} is the covariance matrix between \vec{x} and \vec{z} , and π_x is the covariance matrix

* Research partially supported by NASA under Contract No. NAGW-925.

of \bar{z} . It follows from the multiplicative model that $\bar{\mu}_z = \bar{\mu}_x$, $\pi_{xz} = \pi_x$, and $\pi_x(i, i) = (\sigma_z^2 - \mu_z^2 \sigma_v^2) / (\sigma_v^2 + 1)$, $\pi_x(i, j) = \sigma_{z, i, j}$, for $i \neq j$, $1 \leq i, j \leq M$. In practice, $\bar{\mu}_z$ and π_x are estimated as the sample mean and covariance of the data in a neighborhood of the point to be filtered. The noise power, as will be shown in Sec. 3, is $\sigma_v^2 = N!^2(N)/\Gamma^2(N+1/2) - 1$, where N is number of looks. π_x is then calculated from the above equation. It can be shown that $\sigma_z^2 = \sigma_x^2 + \sigma_v^2 E\{x_i^2\} > \sigma_x^2$, for $1 \leq i \leq M$, while $\sigma_{z, i, j} = \sigma_{x, i, j}$ for $i \neq j$. Thus the correlation between image bands is weakened by the noise; and the effect is much more pronounced than it would be if the noise were additive instead of multiplicative.

Of course, if the terrain reflectances x_i in the separate bands are uncorrelated, i.e. $\sigma_{x, i, j} = 0$ for $i \neq j$, the vector filter reduces to a bank of scalar LMMSE filters [2], [7], $\hat{x}_i = \mu_z + (\sigma_x^2 / \sigma_z^2) (z_i - \mu_z)$, $i = 1, 2, \dots, M$, where $\sigma_x^2 = \pi_x(i, i)$ as given above.

When the terrain reflectances are correlated between bands, the vector filter will result in a better estimate of the terrain reflectance than the scalar filter. To illustrate the point, consider the case $M = 2$. The mean-squared error $MSE_v = E\{|x_1 - \hat{x}_1|^2\}$ in the vector estimate of x_1 is given by

$$MSE_v = \sigma_x^2 - \frac{\sigma_x^4 \sigma_z^2 - 2\sigma_{x_1 x_2}^2 \sigma_z^2 + \sigma_{x_1 x_2}^2 \sigma_z^2}{\sigma_z^2 \sigma_z^2 - \sigma_{x_1 x_2}^2};$$

whereas if we base our estimate of x_1 only on z_1 , then $MSE_s = \sigma_x^2 - \sigma_x^4 / \sigma_z^2$. The difference is $MSE_v - MSE_s = -\sigma_{x_1 x_2}^2 (\sigma_z^2 - \sigma_z^2) / [\sigma_z^2 (\sigma_z^2 - \sigma_{x_1 x_2}^2)] \leq 0$. Figure 1 shows as a function of the correlation, the decrease in mean-squared error that results with the vector filter.

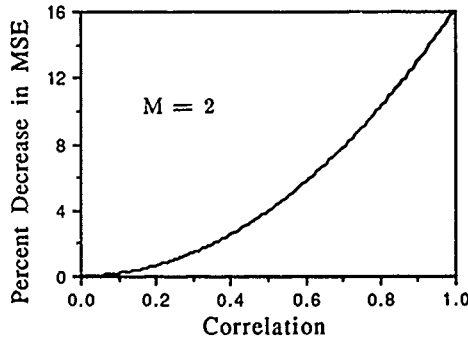


Fig. 1. Decrease in mean-squared error resulting from use of the vector LMMSE filter instead of the scalar LMMSE filter as a function of the correlation between two image bands. The terrain reflectance means and variances are assumed to be $\mu_{x_1} = \mu_{x_2} = 100$, and $\sigma_{x_1}^2 = \sigma_{x_2}^2 = 150$.

Spatially registered images acquired over the same terrain region via modalities other than SAR, may also be used to reduce speckle noise in the SAR images. We used a thematic mapper (TM) image for this purpose. In this case, the noise is modeled as being additive; so assuming that the i -th band contains the TM image, $\pi_x(i, i) = \sigma_z^2 - \sigma_v^2$, where σ_v^2 is again the noise power. Nothing else in the analysis changes. Since the TM images are much less noisy than the SAR images, a more significant decrease in MSE is obtained.

3. MAXIMUM LIKELIHOOD CLASSIFICATION BASED ON A MULTIPLICATIVE NOISE MODEL

Speckle noise results from the interference of wavelets from different scatterers on a rough surface. Using the random walk model, it was shown that the intensity of the

speckle noise is exponentially distributed [9]. To reduce speckle noise, the SAR images are preprocessed by averaging N looks on an intensity basis, and then computing amplitude. The average of N independent exponentially distributed random variables has an N -th order Erlang distribution

$$f_x(x) = \alpha^N / (N-1)! x^{N-1} e^{-\alpha x}, \quad x \geq 0.$$

Hence, the amplitude of the speckle noise has a square-root-of-gamma distribution

$$f_v(v) = 2\alpha^N / (N-1)! v^{2N-1} e^{-\alpha v^2}, \quad v \geq 0,$$

with mean $\mu_v = \Gamma(N+1/2) / [\alpha^{1/2} \Gamma(N-1)]$, and variance $\sigma_v^2 = N/\alpha - [\Gamma(N+1/2) / \Gamma(N-1)]^2 / \alpha$. Under this model, the speckle index [9] for a 4-look SAR image is $\sigma_v / \mu_v = 0.2535$. In Fig. 2, the square-root-of-gamma probability density function with $\mu_v = 1$ and $N = 4$ is plotted, together with the Gaussian probability density function with $\mu_v = 1$ and $\sigma_v = 0.2535$. The curves are very close indeed, indicating that to a good approximation, the speckle noise can be assumed to have a Gaussian distribution.

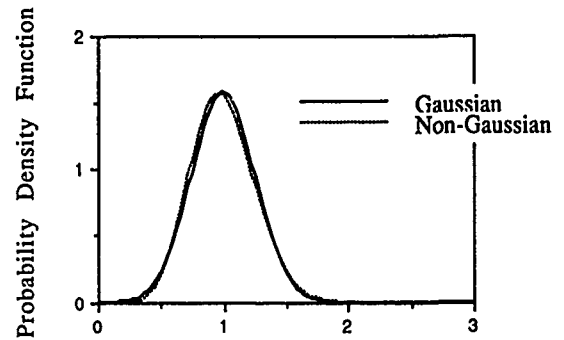


Fig. 2. The square-root-of-gamma probability density function with $\mu_v = 1$, and $N = 4$, compared to the Gaussian probability density function with $\mu_v = 1$ and $\sigma_v = 0.2535$.

To classify an N -look SAR image, we estimate class statistics from training samples, and then design the ML classifier. Assume that there are N_c classes each containing M_i training samples, $i = 1, 2, \dots, N_c$. Under the multiplicative noise model, the k -th sample $z_{k,i}$ of the i -th class can be represented as $z_{k,i} = x_i v_{k,i}$, where x_i is the terrain reflectance of the i -th class, and $v_{k,i}$ is the speckle noise for the k -th data point of the i -th class. Assuming a square-root-of-gamma distribution for the speckle noise, we obtain a ML estimate for x_i of the form

$$x_i = \left[\frac{\alpha}{M_i N} \sum_{k=1}^{M_i} z_{k,i} \right]^{1/2},$$

where $\alpha = [\Gamma(N+1/2) / (N-1)!]^2$ since $\mu_v = 1$. For class i , the discriminant function of data point z_i , defined to be the log-likelihood function, is

$$g_i(z) = (2N-1) \ln(z/\hat{x}_i) - \ln(\hat{x}_i) - \alpha(z/\hat{x}_i)^2,$$

where $i = 1, \dots, N_c$. We assign z to that class for which the discriminant function has the largest value.

If we approximate the noise probability density function by a Gaussian density, then the ML estimate for x_i is

$$\hat{x}_i = \left\{ -\sum_{k=1}^{M_i} z_{k,i} + \left[\left(\sum_{k=1}^{M_i} z_{k,i} \right)^2 + 4M_i \sigma_v^2 \sum_{k=1}^{M_i} z_{k,i}^2 \right] \right\} / 2M_i \sigma_v^2;$$

and the discriminant function for data point z is

$$g_i(z) = -\ln(\hat{x}_i) - (z - \hat{x}_i)^2 / (2\sigma_v^2 \hat{x}_i^2).$$

4. SEQUENTIAL CLASSIFICATION

By considering the values of pixels z_1, \dots, z_n in a neighborhood of the pixel z_0 to be classified, we can increase classification accuracy over that of a per-pixel classifier, which bases its decision about z_0 only on the value of z_0 . Let $\bar{z} = (z_0, \dots, z_n)$ denote the data sequence. For each class i , the generalized sequential probability ratio [10] is computed according to

$$L_i^n(\bar{z}) = f_i^n(\bar{z}) / \left(\prod_{j=1}^{N_c} f_j^n(\bar{z}) \right)^{1/N_c},$$

where $f_i^n(\bar{z}) = \prod_{k=0}^n f_i(z_k)$ is the joint probability density of the data, assuming that it belongs to class i . The test is to compare $L_i^n(\bar{z})$ with a threshold a_i . If $L_i^n(\bar{z}) < a_i$, then z does not belong to class i ; and we are left with $N_c - 1$ classes. We add another sample point to the data sequence, and repeat the test with $N_c - 1$ classes. When only one class is left, the data point z_0 is assigned to it. The thresholds are given by $a_i = (1 - c_i) / \left(\prod_{j=1}^{N_c} (1 - c_j) \right)^{1/N_c}$, where c_j is the probability of assigning z_0 to class i when it actually belongs to class j . These probabilities are specified as part of the classifier design.

5. ROBUST SEQUENTIAL CLASSIFICATION

Consider the problem of assigning a sample z to one of 2 classes with probability densities

$$f_i(z) = (1 - \epsilon)f_i^N(z) + \epsilon f_i^D(z), \quad i = 0, 1;$$

where $f_i^N(z)$ is the known nominal density function and $f_i^D(z)$ is an unknown density function which accounts for deviations from the assumed model. To limit the effect of these deviations on the classification decision, we form the robust likelihood ratio [11]

$$l(z) = \begin{cases} c_0, & l^N(z) \geq c_0, \\ l^N(z), & c_1 < l^N(z) < c_0, \\ c_1, & l^N(z) \leq c_1, \end{cases}$$

where $l^N(z) = f_1^N(z) / f_0^N(z)$ is the nominal likelihood ratio, and c_0 and c_1 are solved from

$$P_0^N[l^N(z) < c_0] + 1/c_0 P_1^N[l^N(z) \geq c_0] = 1/(1 - \epsilon),$$

$$P_1^N[l^N(z) > c_1] + c_1 P_0^N[l^N(z) \leq c_1] = 1/(1 - \epsilon),$$

using Newton's method. $P_i^N[A]$ denotes the probability of event A under the nominal distribution for z , assuming that z belongs to class i . To apply the robust likelihood ratio to M -ary sequential classification, we first use the generalized sequential probability ratio test to select the two most probable classes; and then make the final decision using $l(z)$.

6. EXPERIMENTAL RESULTS

The images used in the experiment were obtained by the Shuttle Imaging Radar-B (SIR-B) during Space Shuttle Flight 41-G in October, 1984. Three sets of data were collected over a forested area in northern Florida on Oct. 9, 10, and 11, at incidence angles 58° , 45° , and 28° , respectively. The images were digitally processed by the Jet Propulsion Laboratory (JPL) in Pasadena, CA. A set of Thematic Mapper (TM) images over the same area was obtained on Oct. 12, 1984; and they were registered with the SIR-B images by JPL. The images we used here have 512×512 pixels. Each pixel covers a $28.5 \times 28.5 \text{ m}^2$ area on the ground. To reduce speckle noise, the images have been preprocessed by averaging the intensity of four looks, then computing amplitude. Sixty sample fields with known ground truth were selected, out of which 18 were training fields, and 42 were test fields. These sample fields represent six ground cover classes: water, clear-cut and pine (1 - 3 years), pine (6 - 36

years), pine (> 36 years), mixed swamp, and cypress swamp.

The three SIR-B image bands were filtered together with the vector filter, and separately with the scalar filter. Each SIR-B image band and band 5 of the TM image were also filtered with the vector filter. The size of the window within which sample statistics were computed was 5×5 . It was observed that the estimates of $\sigma_{x_i}^2$ were sometimes negative. This mainly occurred in two kinds of regions: those where the terrain reflectance fluctuated very little, so that $\sigma_{x_i}^2 \approx 0$; and those where many pixel values were 255 due to clipping. We modified the algorithms so that whenever $\sigma_{x_i}^2$ is negative, it is forced to be zero. In this case, the estimate of the terrain reflectance will be the local mean. Figure 3 shows portions of the images filtered with both the unmodified and modified filters. It can be seen that many artifacts are removed by our modification of the filter, while linear features are almost unaffected.

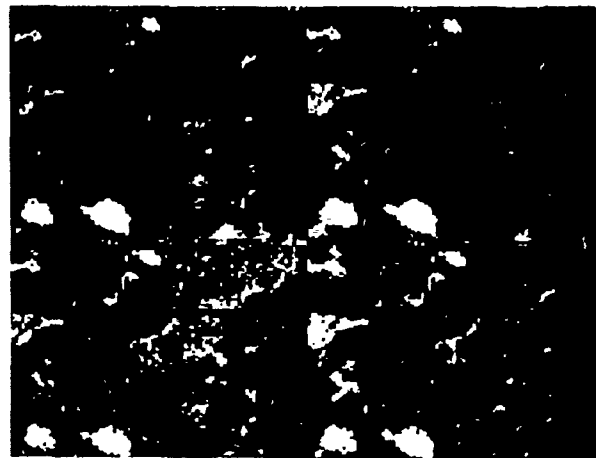


Fig.3. Effect of different processing techniques on a 3 band SIR-B SAR image: unmodified scalar filter, modified scalar filter, unmodified vector filter, and modified vector filter (left to right and top to bottom).

To further illustrate this point, we ran a simulation. Three-variate Gaussian random data x_1, x_2 , and x_3 were generated with parameters $\mu_{x_i} = 100$, $\sigma_{x_i}^2 = 150$, and $\sigma_{x_i x_j} = 75$ for $i, j = 1, 2, 3$. These were multiplied separately by uncorrelated Gaussian random variables with $\sigma_v = 0.25$ to obtain three bands of simulated images z_1, z_2 , and z_3 . We first processed the simulated images with scalar and vector filters based on the true statistics $\mu_{x_i}, \sigma_{x_i}^2$, and $\sigma_{x_i x_j}^2$, and calculated the mean-squared error in the estimate of x_1 . The MSE for the scalar filter was 121.2, and that for the vector filter was 112.7. We then repeated the processing with unmodified filters using sample statistics to estimate $\mu_{x_i}, \sigma_{x_i}^2$, and $\sigma_{x_i x_j}^2$. In this case, the MSE for the scalar filter increased to 189.2, and that for the vector filter increased to 243.1. Modifying the filter algorithms to force negative sample variances to zero reduced the MSE to 164.4 for the scalar filter, and 189.5 for the vector filter. The MSE for the vector filter is larger than that for the scalar filter in both cases where sample statistics are used. This is due to the fact that for a fixed sample size, the estimate of the covariance matrix becomes less accurate at larger dimensions [12]. Increasing the window size to 11×11 reduced the MSE for the scalar filter to 133.6, and that for vector filter to 134.6. At a window size of 17×17 , the MSE's were 127.2 and 122.5, respectively. These results are summarized in Table 1.

Table 1. Mean-Squared Error in Filter Output for Different Parameter Estimates

| | Based on True Stat. | Based on Sample Statistics | | | |
|--------|---------------------|----------------------------|-------------|---------------|---------------|
| | | Unmod. 5×5 | Mod. 5×5 | Mod. 11×11 | Mod. 17×17 |
| Scalar | 121.2 | 189.2 | 164.4 | 133.6 | 127.2 |
| Vector | 112.7 | 243.1 | 189.5 | 134.9 | 122.5 |

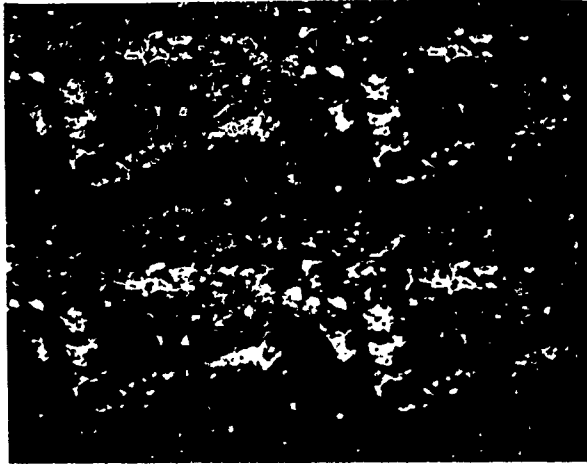


Fig. 4. Effect of different filters on a 3 band SIR-B SAR image: original, scalar filter, vector filtering of 3 SAR bands together, and vector filter applied separately to each SAR band with 1 TM band (left to right and top to bottom).

Figure 4 shows portions of the original image, and images processed by applying scalar filters to each SAR band, a single vector filter to all 3 SAR bands, and separate vector filters to each SAR band with 1 TM band. There are distinct differences in the sharpness of the images processed by the three approaches. Vector filtering of each SAR band in combination with 1 TM band produces the sharpest image followed by vector filtering of the 3 SAR bands together. The scalar filter produces the least sharp image. To gain a more quantitative assessment of how the processing methods differ in their effect on images, we calculated the percent increase in average edge spread and the ratio of average road contrast relative to that of the original image, as defined in [7], and classification performance using a conventional per-pel ML classifier based on the 3 SAR bands. The results are shown in Table 2.

Table 2. Filter Performance

| Filter Type | Scalar | Vector 3 SAR Bands | Vector 1 SAR Band 1 TM Band |
|-------------------------|--------|--------------------------|-----------------------------------|
| Edge Spread Increase(%) | 25.9 | 16.2 | 21.3 |
| Road Contrast Ratio | 0.567 | 0.653 | 0.695 |
| Classification Rate (%) | 84.1 | 82.4 | 83.4 |

In terms of edge spread and road contrast, the ranking of the 3 processing methods agree with their visual appearance. However, the gain in image sharpness is achieved at the cost of a small decrease in classification rate.

For the second part of the experimental work, we examined the performance of the three multiplicative-model-based classifiers discussed in Secs. 3 - 5 on the first band of the original SIR-B SAR image. The three classifiers are per-pel ML, sequential, and robust sequential. All classifiers were evaluated under both the square-root-of-gamma and Gaus-

sian noise distributions. The results are shown in Table 3.

Table 3. Classifier Performance (Per Cent Correct)

| Classifier | Non-Gaussian | Gaussian |
|-------------------|--------------|----------|
| Per-Pel ML | 59.4 | 59.6 |
| Sequential | 78.6 | 78.2 |
| Robust Sequential | 79.9 | 80.8 |

The sequential classifier performs substantially better than the per-pel ML classifier on 1 band of the original image, and almost as well as the per-pel ML classifier applied to 3 bands of filtered images, as indicated in Table 2. The robust procedure yielded a small additional improvement. There was no significant difference in the performance of the classifiers designed under the two different noise distributions. Applying the sequential classifier to 3 bands of filtered images will yield the highest classification rate [7].

7. CONCLUSIONS

Vector linear minimum mean-squared error (LMMSE) filtering can exploit the correlation between bands in a multi-band image to reduce the mean-squared error in the filtered image. However, filter performance may be degraded by poor estimates of the covariance matrix. In our experiments, the vector filter did result in sharper images than the scalar LMMSE filter. Maximum likelihood (ML) classifiers based on a multiplicative noise model have a significantly different and somewhat simpler structure than those based on additive Gaussian noise. However, the two approaches do not result in significantly different classifier performance.

REFERENCES

- [1] L. J. Porcello, *et al.*, "Speckle Reduction in Synthetic Aperture Radars," *J. Opt. Soc. Am.*, Vol. 66, pp. 1305-1311, Nov. 1976.
- [2] J. S. Lee, "Statistical Modelling and Suppression of Speckle in Synthetic Radar Images," *Proc. of IGARSS '87 Symposium*, Ann Arbor, MI, 18-21 May 1987, pp. 1331-1339.
- [3] T. R. Crimmins, "Geometric Filter for Reducing Speckle," *Opt. Eng.*, Vol. 25, pp. 651-654, May 1986.
- [4] J. S. Lim, H. Nawab, "Techniques for Speckle Noise Removal," *Opt. Eng.*, Vol. 20, pp. 472-480, May 1981.
- [5] V. T. Tom, M. J. Carlotto, "Adaptive Least-Squares Technique for Multi-band Image Enhancement," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Tampa, FL, 26-29 March 1985, pp. 704-707.
- [6] N. P. Galatsanos, R. T. Chin, "Digital Restoration of Multichannel Images," *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-37, pp. 415-421, Mar. 1989.
- [7] Q. Lin, J. P. Allebach, "Improving Cover Type Identification in Speckled SAR Images by Prefiltering and Sequential Classification," *Proc. of 26th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, 28-30 Sep. 1988. Also submitted to *IEEE Trans. Geosci. Remote Sensing*.
- [8] J. M. Durand, B. J. Gimonet, and J. R. Perbos, "SAR Data Filtering for Classification," *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-25, pp. 629-637, Sep. 1987.
- [9] J. W. Goodman, "Statistical Properties of Laser Speckle Patterns," in *Laser Speckle and Related Phenomena*, Edited by J. C. Dainty, New York: Springer-Verlag, 1984.
- [10] K. S. Fu, *Sequential Methods in Pattern Recognition and Machine Learning*, New York: Academic Press, 1968.
- [11] H. V. Poor, *An Introduction to Signal Detection and Estimation*, New York: Springer-Verlag, 1988.
- [12] H. M. Kalayeh, D. A. Landgrebe, "Predicting the Required Number of Training Samples," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-5, pp. 664-667, Nov. 1983.

VISUAL EFFECT OF SPECKLE REDUCTION ON INTERPRETATION OF ONE-LOOK SAR PHOTO IMAGES*

Hiroshi Kimura, Natsuhiko Motomura
and Nobuhiko Kodaira

Remote Sensing Technology Center of Japan
7-15-17, Roppongi, Minato-ku,
Tokyo 106, Japan

Haruto Hirose

Institute of Space and Astronautical Science
3-3-1, Yoshinodai, Sagami-hara-shi,
Kanagawa 229, Japan

ABSTRACT

One-look SAR photo image which is printed in an appropriate scale seems to be speckle-reduced in some degree and yet to keep a high spatial resolution. It is thought this reason is that gradation caused by photographic process and interpretation process of human eyes has the equivalent effect to multiple look and works to suppress fluctuation of speckles. This paper describes a discussion about photographic process and visual effect on one-look SAR photo images, and qualitative analysis results with respect to the visual effect.

Keywords: SAR, Speckle, One-look, Photo image, Visual Effect, Photographic process, Multiple look

1. INTRODUCTION

Synthetic aperture radar images have an inherent noise called "speckle". In signal processing, multiple look processing is generally applied to reduce this noise. By N-look multiple look processing, standard deviation of the intensity fluctuation of speckle can be suppressed to $1/\sqrt{N}$, but the spatial resolution becomes N times lower. Standard product of spaceborne SAR, such as SEASAT SAR, SIR-B, EERS-1 and JERS-1, is 3 or 4 look image.

In a part of the research and development project on observation system for JERS-1, authors simulated one, two and three look SAR images and evaluated them with several specialists in the field of radar, data processing and applications. Figure 1. shows an example of the simulated SAR photo images, and Table 1. describes specs of these photo images.

In compression of one-look photo images and 3-look ones, it was found that one-look photo

images don't seem to be as bad as expected, because the speckle doesn't appear remarkably and fine spatial detail is preserved in them. Although the standard deviation of intensity fluctuation of speckle in one-look image is 1.7 times as high as that in 3-look image, photo images don't look different so much. We think that this is due to the gradation caused by photographic process and interpretation process by human eyes. From this point of view, we study an appearance of difference in number of look in SAR photo images.

Table 1. Specs of SAR Photo Images

| No. of Looks | 1 | 3 |
|------------------------------|-----------|-------------|
| Resolution | 6mx18m | 18mx18m |
| Pixel spacing | 5mx5m | 10mx10m |
| Pixel size on film and paper | 50µmx50µm | 100µmx100µm |
| Printing | Contact | Contact |
| Photo image scale | 1:100,000 | 1:100,000 |

2. PRODUCTION AND INTERPRETATION OF PHOTO IMAGE

2.1 Concept

Figure 2 shows concept of production and interpretation of the photo image. Photo images are produced from digital data with a photographic system, including film recorder, film developing, printing and paper developing. These photo images are interpreted in brain through visual system of human eyes. In these two transfer systems, image gradation and brightness change occur,

*This study is a part of the national research and development project on observation system for Earth Resources Satellite-1, conducted under the program set up by the Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

and the degree of them is related to the transfer functions of these two systems.

2.2 Characteristics of Photographic System

Basic four stages of the photographic process are as follows:

- Writing image on film with the recorder,
- Developing the film,
- Printing image on paper from the film,
- Developing the paper.

Transfer function of the whole photographic system is multiplication of transfer functions of each four stages. Four transfer functions depend upon characteristics of the film recorder, film, printing instrument and paper respectively. Gradation which occurs in the photographic system has pixel smoothing effect.

In the case of photo images of Figure 1, resolvable performance of the film is 100 to 200 lines a millimeter, and that of the paper is 50 to 60 lines a millimeter, and contact printing is done. So gradation which occurs at the last three stages can be assumed to be negligible. At the first stage, 50 μ m pixel appears 60 μ m square on the film with the film recorder(OPTRONICS C4300). Gradation caused by this stage has some pixel smoothing effect, but can be thought small in this case.

However, in general, gradation proceeds as enlargement and generation of film increase, for instance second or third generation and so on. So gradation won't be negligible in such case.

2.3 Characteristics of Visual System

Resolvable limitation of human eyes that two adjacent points with similar brightness can be separated into two points depends upon a cell size of visual nerve. Physiological measurement and experiment show that a minimal retinal disparity angle is about 0.00022 radian(Shino,1967). This corresponds to about 75 μ m on the photo in viewing at a distance of 30cm.

When the visual resolution is larger than the pixel size in photo image, each pixels looks blurry, and it's impossible for eyes to resolve the image into pixels. We think that this blur has the same effect as multiple look, and reduces the fluctuation of speckles. The degree of this visual effect depends upon a viewing distance and a scale of photo image. If the pixel size in photo image is smaller than visual resolution, this effect will become higher.

Pixel size of one-look photo image in Figure 1 is 50 μ m, and this is smaller than 75 μ m that is the visual resolution at a viewing distance of 30cm. Therefore, gradation of pixels occurs in some degree for one-look photo image. Pixel size of 3-look photo image is 100 μ m. Because this is larger than the

visual resolution, gradation won't occur in viewing it. We think that this is the reason why one-look image seems good.

3. QUALITATIVE ANALYSIS OF VISUAL EFFECT

To study the visual effect qualitatively, density of photo image was digitized by a densitometer as shown in the right part of Figure 2. Aperture sizes of the densitometer were selected 50 μ m, 100 μ m and 200 μ m. Each sizes correspond to the viewing distances of about 20cm, 40cm and 80cm respectively. Figure 3 shows density histograms of the photo images after digitizing. This figure includes digital value histograms of original image and expected density histograms obtained by using only the relation between digital value and photo density. This relation was measured at the gray scale which was written on the film and printed in the paper together with the image.

- Difference in the photo density histograms between one-look images and 3-look ones is much smaller than that in the expected histograms between them. Expected density histogram is assumed to be the histogram after through the system with no gradation, but density change. Therefore, above result suggests that gradation of the visual system makes difference in appearance between one-look image and 3-look one small.
- At 50 μ m aperture, histograms of one-look photo image spread wider than those of 3-look one. But at 100 μ m aperture, histograms of one-look photo image and those of 3-look one are similar, and they come almost same at 200 μ m aperture. This result corresponds to the fact that one-look photo image and 3-look one seem different in viewing at a short distance such as 20cm, but they look same at a long distance such as 80cm.

From above, we can think an appearance of SAR photo image as follows: In interpretation of the photo image from SAR with around 10m resolution in the scale of 1:100,000, pixels are blurred by the visual effect. This works the same as the multiple look. So difference in appearance of speckle fluctuation among images with different look number becomes smaller, if the viewing distance becomes longer. Change of viewing distance is equal to the change of the photo scale. So if the scale becomes smaller, the difference becomes smaller. Change of the photo scale accompanies the photographic process. In this case, gradation caused by this can't be negligible.

4. CONCLUSION

We have discussed the gradation caused by the photographic process and visual effect with

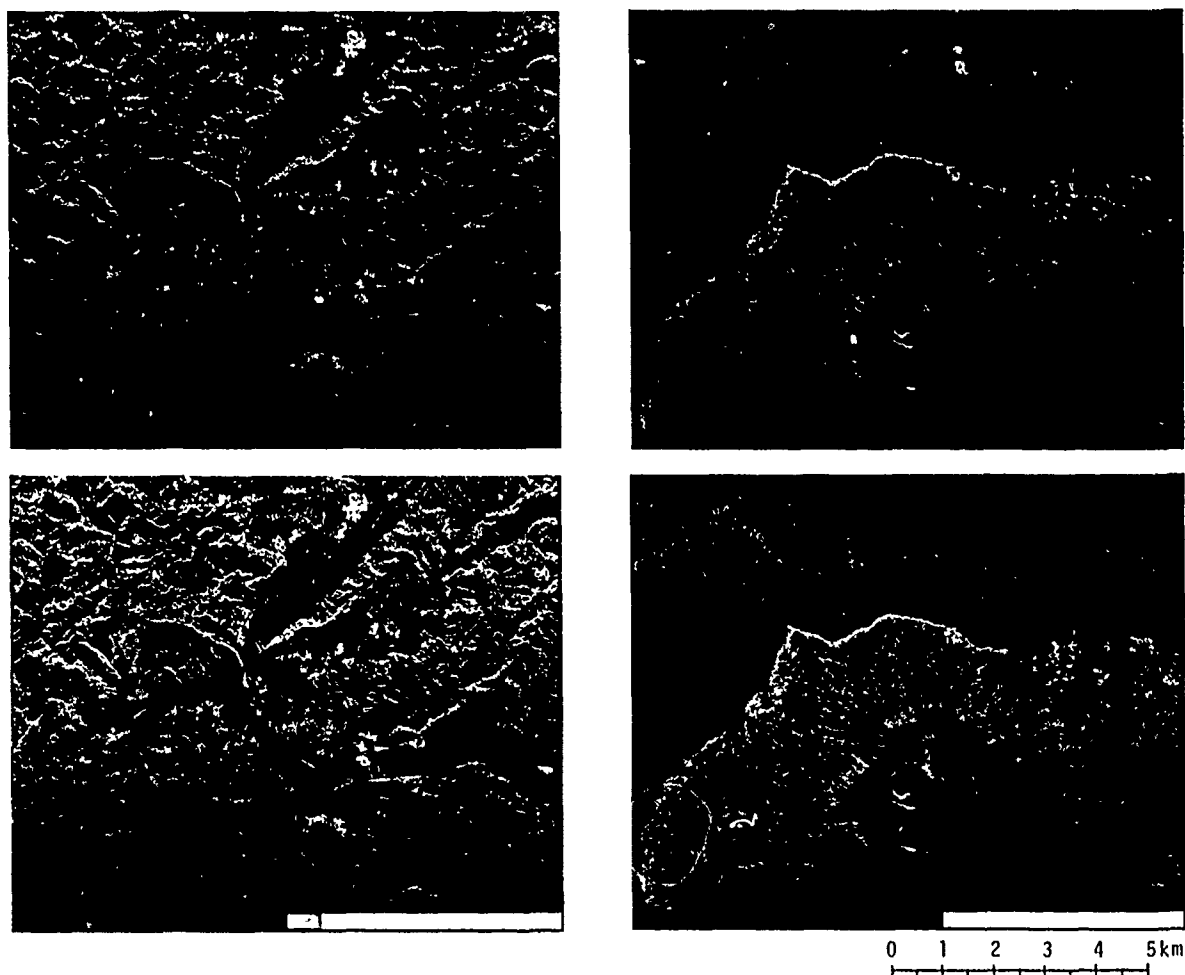


Figure 1. One-look(above) and 3-look(below) SAR photo images. The left is Futatsui and the right is Oshima. Scale of the original copy is 1:100,000.

respect to the appearance of one-look SAR photo images. It is generally considered that multiple look image is better than one-look image. However, the results show that one-look image printed in the proper scale seems to be speckle-reduced and yet reserve the fine spatial resolution. So one-look photo image becomes very useful for interpretation. In the photographic process, gradation proceeds as enlargement and generation of film increase. Therefore, problem about the proper image scale on the film and paper is also suggested.

5. REFERENCES

1. Shino, K., "Photogrammetry", Sankaido, 1967.

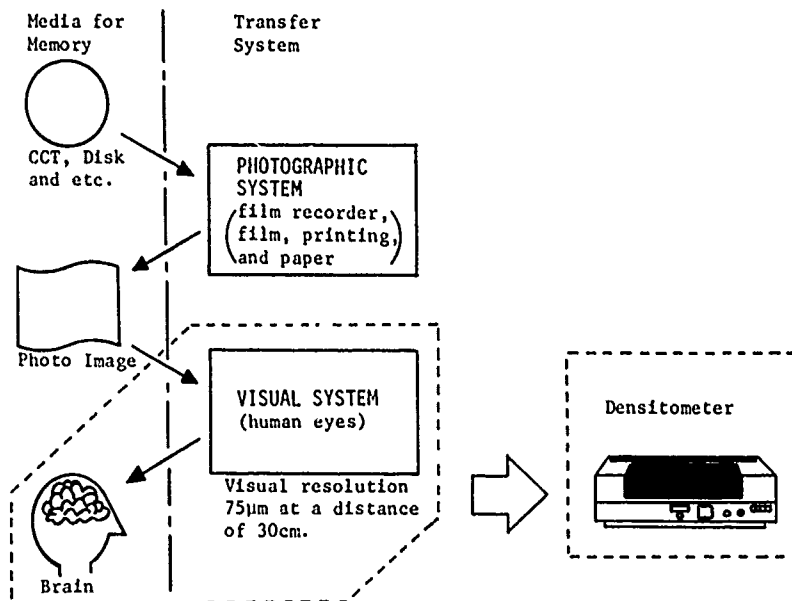


Figure 2. Concept of production and interpretation of the photo image.

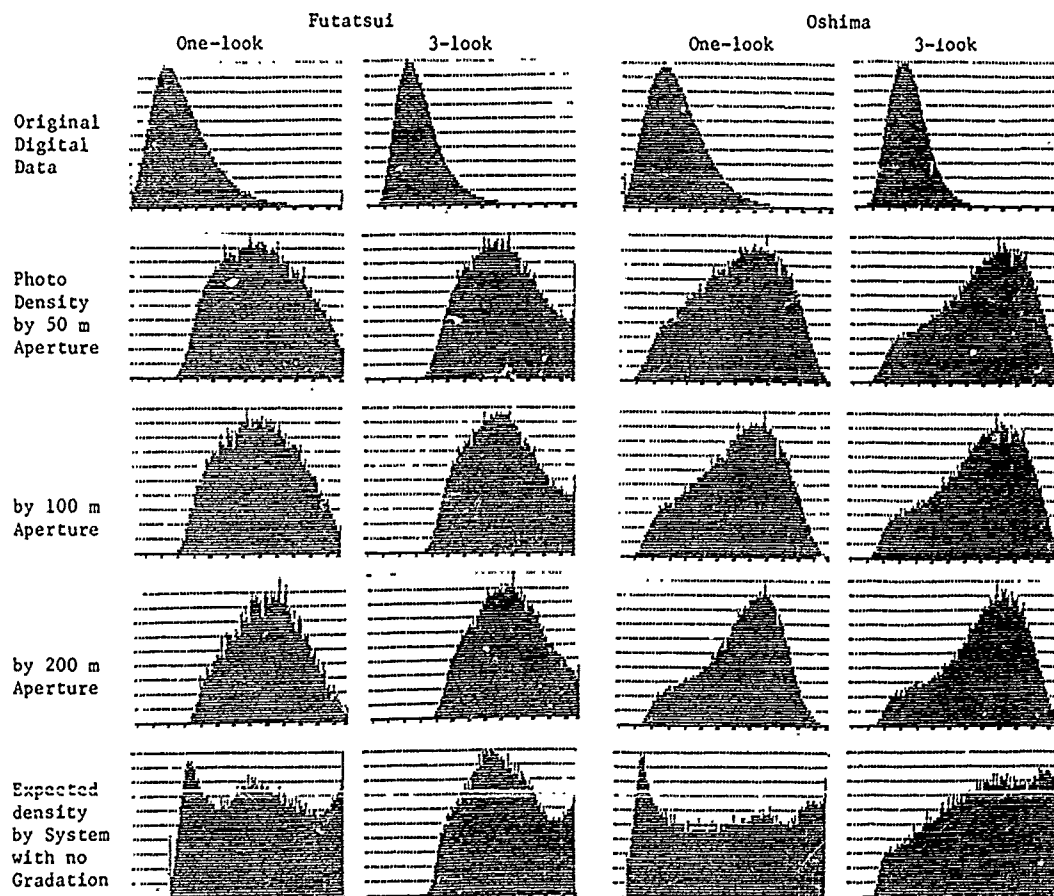


Figure 3. Photo density histograms, including original digital value and expected density histograms.

The FLI airborne imaging spectrometer: experience with land and water targets

J. F. R. Gower,
Institute of Ocean Sciences, P.O. Box 6000, Sidney, BC, V8L 4B2, Canada
Tel: (604) 356-6558, Fax: (604) 356-6479, Telex: 049-7281

R. A. H. Buxton,
Moniteq Ltd., 630 Rivermede Road, Concord, Ont L4K 2H7, Canada
Tel: (416) 669-5334, Fax: (416) 669-3823

G. A. Borstad,
G.A.Borstad Associates Ltd., 9865 West Saanich Rd, Sidney, BC, V8L 3S1 Canada
Tel: (604) 656-5633, Fax: (604) 656-3646

ABSTRACT

An airborne imaging spectrometer, sensitive at visible and near-infrared wavelengths, has been developed by Moniteq Ltd of Toronto for the Canadian Department of Fisheries and Oceans. The instrument has achieved its design goal of imaging the solar-stimulated fluorescence signal from near-surface phytoplankton, and providing the data needed to assess detectability limits and environmental effects on the signal. It has also been used to acquire imagery over a wide variety of water and vegetated land targets. This paper reviews the applications of the instrument from the first five years of its operation in different parts of the world, and shows examples of the imagery.

Key words: Imaging spectroscopy, chlorophyll fluorescence, airborne remote sensing, plant stress.

INSTRUMENT DESIGN

The instrument evolved from the requirement of the Canadian Department of Fisheries and Oceans (DFO) for an improved sensor for mapping phytoplankton distributions. In the design selected, two-dimensional array detectors in CCD (Charge-Coupled Device) cameras built by Itrics Ltd of Calgary are used to give the high spectral resolution and sensitivity required to image solar stimulated in-vivo chlorophyll fluorescence (Neville and Gower 1977, Gower and Borstad 1981). The instrument was therefore named the Fluorescence Line Imager (FLI). For this purpose, spectral resolution of a few nm and signal-to-noise ratios up to 2000 to 1 are required in the spectral range 400 to 800nm.

The FLI was produced by Moniteq Ltd of Toronto in a development programme funded jointly by DFO and the Canadian space programme, in the period 1981 to 1984. The FLI design (Borstad et al. 1985) is based on five CCD arrays (EEV P8600) each containing 385 by 288 sensing elements, each illuminated through a transmission grating imaging spectrometer, and each covering a 14 degree field of view. The full instrument allows collection of high spatial resolution push-broom images with 1900 elements covering a 70 degree swath. Image data are collected in eight spectral bands that can be configured in steps of 1.3 nm. A distinct advantage is that the spectral bandset can be changed easily in flight. An alternative CCD readout mode gives full spectra covering 430 to 805 nm with a resolution of 2.5 nm, but with lower spatial resolution corresponding to 40 image samples across the same 70 degree swath.

The FLI instrument consists of a multiple camera head, controller, data recorder and cooling unit. It has been installed in a variety of aircraft, including the CCRS DC3 and Falcon

Fanjet aircraft, and a Piper Navajo Chieftain (Figure 1).

The FLI has now been used to produce images in a variety of spectral band combinations over targets on land and water (Borstad et al. 1985, Gower et al. 1987, Rock et al. 1988, Buxton 1988, Gower and Borstad 1988). In addition to phytoplankton mapping, its high spectral resolution has also been applied to water depth and submerged vegetation surveys along shorelines. Over land it has been applied to studies of the detailed spectral properties of the chlorophyll red edge for detecting the effects of plant stress. Other uses are in hydrocarbon exploration, landfill site monitoring, agriculture and atmospheric observations.

These studies make use of the FLI's special capability for producing both radiance spectra and images of a given target. Spectral bands for the images can be configured on the basis of the observed spectra.



Figure 1. The FLI imaging spectrometer installed in a Piper Navajo Chieftain aircraft.

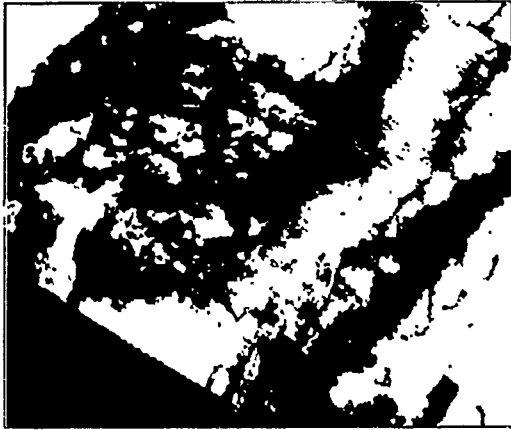


Figure 2. A mosaic image of Barkley Sound B.C., showing solar-stimulated chlorophyll fluorescence on April 25 1987. A digital land mask has been superimposed in white.



Figure 3. An area of high chlorophyll fluorescence contrast, about 4 km square, from the data illustrated in Figure 2. Chlorophyll concentrations are in the range 1 to 4 mg.m⁻³.

In general, compared to other instruments now in use or in development for remote sensing, such as AVIRIS or MODIS, the FLI covers a more limited spectral range, but has higher spectral resolution and sensitivity.

CHLOROPHYLL FLUORESCENCE IMAGING

In a typical example of its application in coastal studies, the FLI was used in April 1987 in support of DFO's MASS (Marine Survival of Salmon) programme to provide aerial coverage of Barkley Sound (48° 50'N, 125° 20'W) on the west coast of Canada. The data have been used to study phytoplankton concentrations, the distribution of fronts and the effects of mixing between water masses in the Sound, at a time when young salmon leave the hatcheries, lakes and rivers in this area and migrate into the open sea.

Figure 2 shows a mosaic image of chlorophyll fluorescence emission stimulated by sun and sky-light for most of the area of Barkley Sound. The mosaic was formed from imagery taken on 6 flight lines at an altitude of 5500m, and covers an area about 20km square. The image data were constructed from three of the eight spectral bands of the FLI (Gower and Borstad 1988). These bands were located at 659 - 673 nm, shortward of the fluorescence region, 673 - 687 nm, which is set to cover the chlorophyll fluorescence emission centred at 685 nm, while avoiding the absorption band due to atmospheric oxygen at wavelengths longer than 687 nm, and 713 - 718 nm where atmospheric absorption is again at a minimum. The grey levels of Figure 2 correspond to fluorescence emission from chlorophyll concentrations in the range 0.5 to 5 mg.m⁻³.

The narrow-band fluorescence signal was found to be relatively insensitive to broader-band atmospheric and surface reflectance effects, making it a useful remote sensing indicator of phytoplankton concentrations.

Figure 3 shows a smaller area of the sound about 4 km square with a spatial resolution of about 20m, where phytoplankton contrast was particularly strong. The small scale structure characteristic of advected phytoplankton is visible in the fluorescence signal but not at adjacent wavelengths where this signal is absent.

An upper limit to the noise level of the fluorescence observations can be estimated from the radiance differences between adjacent pixels in Figure 3. These are independent measurements with the FLI. Some of the differences will be due to real fluorescence variations. The deduced r.m.s. variation of each pixel is about 0.003 W/m².sr.μm, equivalent to about 0.1 to 0.2 mg.m⁻³.

With images similar to Figures 2 and 3, the FLI has demonstrated that remote imaging of solar stimulated chlorophyll fluorescence is indeed possible. The narrow-band fluorescence signal was found useful for mapping phytoplankton concentrations. Also, the fluorescence emission is closely linked to cell physiology, and can vary with nutrient status, species composition and growth rate. In some studies, the fluorescence yield has been shown to be inversely related to the rate at which phytoplankton biomass is formed. This is an essential parameter in the study of ocean and coastal ecosystems.

In general it appears that remote sensing of chlorophyll fluorescence will represent a powerful new tool for aquatic studies.

OTHER APPLICATIONS OVER WATER

The high sensitivity and spectral resolution of the FLI make it also useful for measurements of shallow water bathymetry and for benthic plant surveys. For bathymetric measurements (O'Neill et al., 1986) spectral bands were clustered in the wavelength range of maximum coastal water penetration (500 - 600nm), together with a longer wave band to define atmospheric path radiance, and a shorter wave band to measure chlorophyll concentration by the "green to blue ratio" method. An example of a resulting bathymetric image is shown in Figure 4. Here the imagery has been geometrically corrected using the output from an inertial navigation system, so that image data can be compared with existing charts. Comparison indicates differences of 0.5m in shallow water, increasing to 1.5m in depths greater than 5m.



Figure 4 Geocoded water depth image for depths in the range 0-3m, constructed from FLI output.



Figure 6. A comparison of spectral and imaging mode data of a forested region. Spectral mode data was collected at a lower altitude.



Figure 5. Imaging mode data of the Albion-Scipio oil-field in Michigan USA.

For benthic plant resource assessment, the spectral bands can be set to cover a wider spectral range which includes both the 500 - 600nm water penetration region and the "red edge" region, 690 to 740nm, where chlorophyll reflectance rises steeply with increasing wavelength. An experimental survey of a laminaria (kelp) harvesting area near Yarmouth, Nova Scotia, on the east coast of Canada was carried out with the FLI in July 1985 (Mouchot et al., 1987). In a comparison with aerial photography, FLI imagery was found capable of identifying laminaria growing in deeper water, and of extending the depth range surveyed to beyond the limit of maximum laminaria growth.

APPLICATIONS OVER LAND

Over land, applications have concentrated on the properties of the chlorophyll red edge, which is the dominant spectral feature in the visible reflectance spectrum of plants. The imaging mode of the FLI produces high quality imagery whose bands can be selected as described above. Figure 5 shows a section of an image taken over the Albion-Scipio oil-field in Michigan USA as part of a survey of plant stress due to hydrocarbon seepages.

Operation in the spectral mode give full spectra with 2.5nm resolution over the optical range 430 to 805nm. The spectra are collected as 288 band push-broom imagery from 40 different look directions, giving lower spatial resolution imagery as shown in Figure 6.

Since the spectrum of each pixel of the spectral mode data is available, detailed comparisons can be made of the spectral reflectance of different areas in a study region. The areas can be selected from the low-resolution spectral mode imagery, and regions of the spectrum of special interest, for example that including the chlorophyll red edge, can be extracted as illustrated in Figure 7. This feature is known to exhibit variations in shape and in mean wavelength position, when vegetation is stressed by abnormal environmental conditions. These conditions include atmospheric pollution, precipitation pH, metal, hydrocarbon and toxic substance concentrations in the soil, and climatic variations. Special image processing has been applied to FLI imagery to produce images whose brightness values correspond to red edge position (Figure 7).

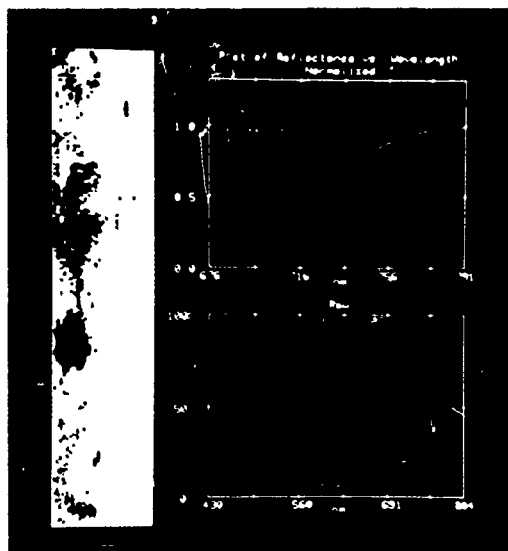


Figure 7. A stress image derived from data collected in the spectral mode of the FLI. The upper spectral curves are model fits to the data indicating a blue shift in an area of known geochemical anomaly.

A comparison of the vegetation stress parameters derived from FLI imagery and spectral mode data from an oil field adjacent to Albion-Scipio showed an encouraging correlation with soil gas measurements (Reid et al., 1988). Similar effects have been observed at mineralized sites due to high metal concentrations. These results show the potential of the technique as an exploration tool.

CONCLUSIONS

FLI data are currently commercially available through Moniteq Ltd, who operate the instrument on loan from DFO. For smaller scale surveys, a more compact and flexible CCD-based imaging spectrometer built by Itres Research Ltd is now also available. This is the Compact Airborne Spectral Imager (CASI), which has evolved from experience with the FLI.

Both the FLI and the CASI have higher spectral resolution and sensitivity, though more restricted spectral coverage, than past or planned (HIRIS, MODIS or SeaWiFS) satellite instruments. The evidence presented here shows that the potential of remote optical measurements of the earth's surface is far from being exhausted with the present generation of space instruments. We look forward to the launch of more advanced sensors in the future.

REFERENCES

- Borstad, G.A., H.R. Edel, J.F.R. Gower, and A.B. Hollinger, 1985, Analysis of test and flight data from the Fluorescence Line Imager, Canadian Special Publication of Fisheries and Aquatic Sciences, No. 83, 38pp.
- Buxton, R.A.H., 1988, "The FLI airborne imaging spectrometer: a highly versatile sensor for many applications", Proceedings, ESA Workshop on Imaging Spectrometry, ESRIN, Frascati, ESA SP-1101, 11-16.
- Gower, J.F.R. and G.A. Borstad, 1981, "Use of the in-vivo fluorescence line at 685 nm for remote sensing surveys of chlorophyll a", in J.F.R. Gower (Ed) "Oceanography from Space" Plenum Press, 329-338.
- Gower, J.F.R., G.A. Borstad, L.H. Gray and H.R. Edel, 1987, "The Fluorescence Line Imager: imaging spectroscopy over water and land", Proceedings of the 11th Canadian Symposium on Remote Sensing, Waterloo, Ontario, June 22-25 1987, 689-697.
- Gower, J.F.R., and G.A. Borstad, 1988, "Mapping of phytoplankton fluorescence with the FLI imaging spectrometer", Proceedings of IGARSS'88 Symposium, Edinburgh, Scotland, Sept 13-16 1988, ESA SP-284, pp1391-1394.
- Mouchot, M.C., G. Sharp and E. Lambert, 1988, "L'utilisation du 'Fluorescence Line Imager' (FLI) pour la cartographie thématique des végétaux marins submergés", Proceedings of the 11th Canadian Symposium on Remote Sensing, Waterloo, Ontario, June 22-25 1987, 699-708.
- Neville R.A., and J.F.R. Gower, 1977, "Passive remote sensing of phytoplankton via chlorophyll a fluorescence", Journal of Geophysical Research, 82, 3487-3493.
- O'Neill, N.T., A.R. Kalinauskis, J.D. Dunlop, A.B. Hollinger, H. Edel, M. Casey and J. Gibson, 1986 "Bathymetric analysis of geometrically corrected imagery data collected using a two-dimensional imager" Proceedings of the Society of Photo-Optical Instrumentation Engineers, 637, 26-33.
- Reid, N.J., A. Iwashita, Y. Yamashita and K. Thompson, 1988, "High resolution imaging of geobotanical anomalies associated with subsurface hydrocarbons" Proceedings of the Sixth Thematic Conference on Remote Sensing for Exploration Geology, Houston, Texas, May, ERIM pp 213-223.
- Rock, B.N., T. Hoshizaki, and J.R. Miller, 1988, "Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline" Remote Sensing of Environment, 24, 109-127.

A COMPACT AIRBORNE SPECTROGRAPHIC IMAGER (CASI)

S.K. BABEY AND C.D. ANGER

ITRES RESEARCH LTD.
110, 6815 - 8TH ST. NE
CALGARY, ALBERTA
CANADA T2E 7H7

ABSTRACT

In this paper we describe a new instrument, the Compact Airborne Spectrographic Imager (CASI), which is capable of operating in light aircraft and acquiring visible and near infrared imagery in digital form. The instrument consists of the sensor head, Instrument Control Unit and Power Supply Module, together with a monitor and keyboard. An f/4 reflection grating spectrograph is coupled to a CCD area sensor, providing a resolution of 578 spatial pixels; the spectrum from 450-950 nm for each spatial pixel is dispersed across 288 spectral pixels. To achieve manageable data rates, two operating modes are defined. Multispectral Imaging mode, which provides maximum spatial resolution for a limited number of spectral bands, and Multispectrometer mode, in which full spectra are collected for a limited number of points in the scene. Flight data may be played back on the instrument itself, or on a PC for calibration and transcription to files for image processing.

Keywords: imaging spectroscopy
airborne instrumentation
multispectral imaging

- (1) High performance VNIR pushbroom imaging spectrograph (high sensitivity, high resolution, wide dynamic range).
- (2) Compact and easily transported (checked luggage).
- (3) Easily installed in small planes (aerial camera mount).
- (4) Simple to operate (non-expert).
- (5) Low Cost (instrument, operation and processing).

Instrument Description

A block diagram of the instrument is shown in Figure 1. The spectrograph and CCD detector are located in the Sensor Head, while data handling and control are performed by the Instrument Control Unit (ICU). The operator interacts with the ICU via a keyboard and monitor. The Power Supply Module (PSM) generates the necessary voltages for the instrument.

The Sensor Head is normally installed with the objective lens oriented downwards such that a line across the flight path is imaged onto the spectrograph slit (Figure 2). The field of view can be selected between 15° and 60° by substitution of different fore-optic lenses. Light emerging from the slit is collimated and dispersed by the reflection grating. The beam

Introduction

Throughout the world, multispectral remote sensing is becoming increasingly important for obtaining resource inventories, performing environmental monitoring, and aiding in the planning of land and resource usage. At the same time, the accelerating demand for data in digital form has prompted interest in instruments capable of directly recording digital imagery. In particular imaging spectrographs such as the Fluorescence Line Imager (Gower, this proceedings) have appeal because of their total flexibility in the choice of wavelength and bandwidth of spectral bands.

Despite their advantages over filter-based scanners, however, the use of airborne imaging spectrographs has been limited, primarily due to the fact that there are few instruments available, and the cost of building and flying those instruments is relatively high.

In order to overcome these difficulties, a new instrument, the Compact Airborne Spectrographic Imager (CASI) has been developed. The CASI was designed with the following goals in mind:

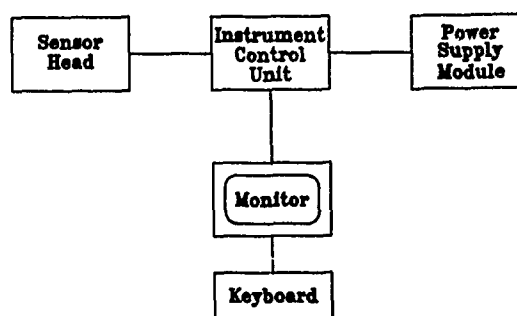


Figure 1 Block Diagram of the Instrument

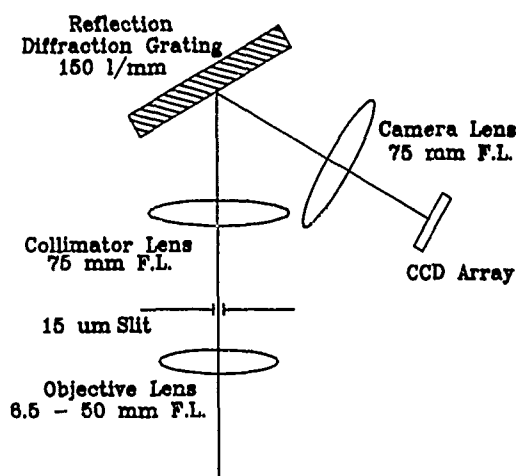


Figure 2 CASI Optical System

is then focused onto the CCD. There is a motorized Iris that can be used to set the focal ratio at $f/4$ or $f/11$.

The sensor is a thermoelectrically cooled P86520 series frame transfer device manufactured by EEV Inc. (Chelmsford, UK). The chip image area contains 578×288 pixels (each $15 \times 22 \mu$). The format is shown in Figure 3. The CCD is oriented to obtain 578 pixels of spatial resolution across the flight path. The spectrum is dispersed along the columns of the CCD, providing 288 spectral resolution elements (each 1.8 nm wide) and covering the spectral range from 450-950 nm. Alternative spectral ranges can be obtained by changing the grating and/or its angle. Successive sensor readouts result in a two-dimensional multispectral image along the flight path.

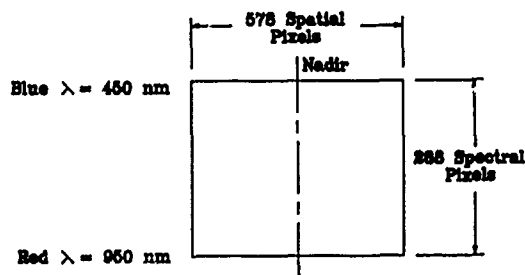


Figure 3 CCD Detector Format

In principle one would like to record all of the pixels in each CCD frame. However, this results in a relatively large quantity and rate of data. For example, if a modest scan rate of 50 frames/sec is assumed, the required throughput would be:

$$\text{Data Rate} = 578 \times 288 \text{ pixels/frame} \times 50 \text{ frames/sec} \times 2 \text{ bytes/pixel} = 16.6 \text{ MB/sec}$$

A one hour flight would produce 60 Gigabytes of data! At the present time, this quantity and rate of data (both recording and analysis) is inconsistent with a low cost philosophy.

Clearly, selectivity is the key to obtaining manageable data rates. The CASI provides two distinct modes of sensor readout to effect this. In Multispectral Imaging Mode, the instrument can be programmed to sum the data from adjacent rows on the CCD Sensor in order to form arbitrary but nonoverlapping spectral bands (Figure 4). This provides the ability to obtain maximum spatial resolution (578 pixels) for a limited number of spectral bands.

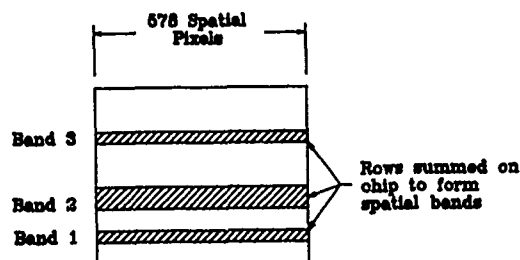


Figure 4 Multispectral Imaging Mode Operation

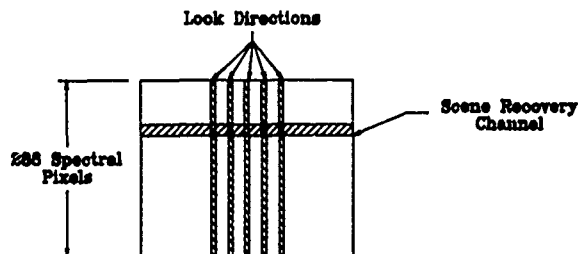


Figure 5 Multispectrometer Mode Operation

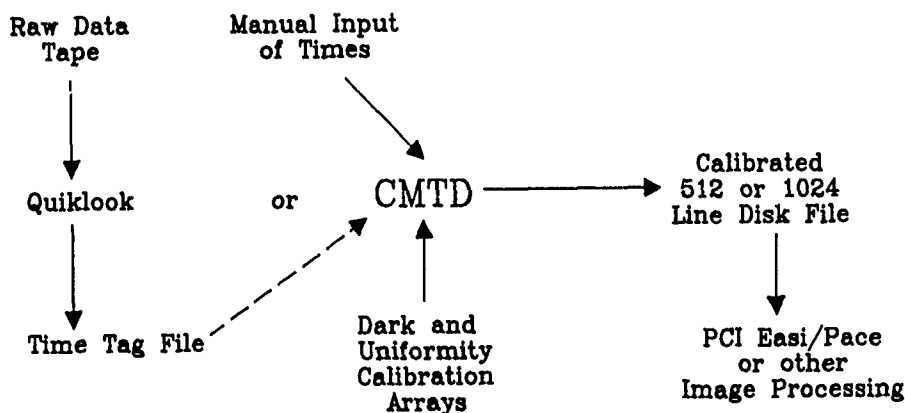


Figure 6 Data Processing Sequence

In Multispectrometer Mode, the reverse is true; that is, all 288 spectral elements are collected for each of a limited number of "look directions" (Figure 5). The number and spacing of the look directions is programmable. It has also been found convenient to collect a full scene image at one wavelength (as in Imaging Mode) in order to permit correlation of the spectral data with features in the scene. Multispectrometer mode is useful for algorithm development and other situations where detailed spectra are of interest, and where reduced spatial resolution can be tolerated.

The mode and configuration are defined by the operator prior to the start of data collection through a menu-driven user interface (configurations may be saved and retrieved). A real-time display presents a pseudocolour or grey-scale image of any band, and signal levels are monitored graphically. Data are recorded on an Exabyte EXB-8200 digital helical scan tape recorder. This unit provides 2 GB capacity on a single 8mm video cassette. The maximum record rate is about 200 KB/sec. This results in a maximum scan rate of 100 lines/sec for a single band. For multiple bands the achievable scan rate is reduced accordingly.

Data Analysis

A useful feature is the QUIKLOOK program, which provides the ability to play back data tapes in order to "re-fly" flight lines. This program can be run on the instrument itself in the field; it can also run on a PC equipped with an Exabyte drive as part of the data extraction and calibration process, shown in Figure 6. The QUIKLOOK program is capable of producing 'time-tag' files containing start and end times for segments of interest. The CASI Magtape To Disk (CMTD) program takes the time tag file (or manually entered times) as input, reads back the recorded tapes and produces disk files for selected bands (Imaging Mode) or look directions (Multispectrometer Mode). The disk files are presently generated in a format compatible with the EASI/PACE image processing package produced by PCI Inc. The CMTD program optionally performs dark current removal and radiometric calibration on the raw data.

Figure 7 shows an uncorrected image of the confluence of the Fraser and Stave Rivers obtained by the CASI on Feb. 8, 1989 from an altitude of 1500 m. Smoke from the sawmill is visible, as are shore ice, log booms and bridge pylons.

Conclusion

A summary of the instrument specification is given in Table 1. The CASI represents a powerful new tool for the multispectral remote sensing community. Plans for future enhancements to the instrument include an irradiance sensor to measure the downwelling flux (this will simplify computation of reflectance values), incorporation of attitude and navigation data (to enable resampling for input to Geographic Information Systems), and extension of spectral sensitivity into the infrared region.

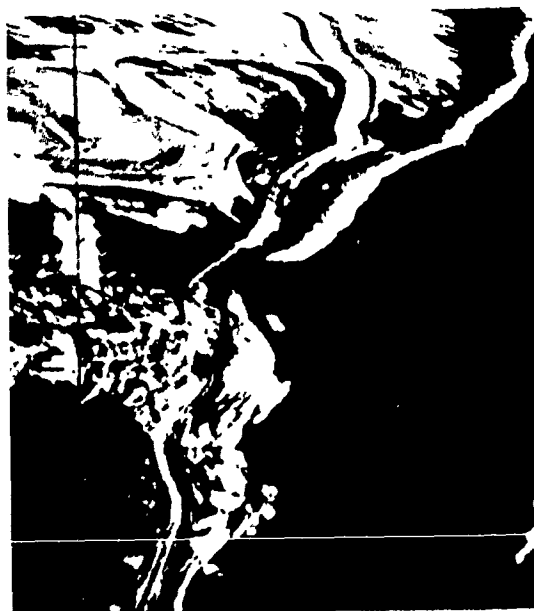


Figure 7

Sample imagery of the Fraser River obtained Feb. 8, 1989 at an altitude of 1500 m. The image is a composite of Band 2 (705-712nm), Band 3 (632-642 nm) and Band 4 (465-477 nm).

Table I

Summary of Instrument Specifications

| | |
|---------------------|--|
| Swath (FOV) | 15° - 60° |
| Spectral Range | 450 - 950 nm |
| Spatial Resolution | 578 pixels |
| Spectral Resolution | 288 pixels (1.8 nm each) |
| Aperture | f/4 or f/11 |
| Sensitivity | 0.08 $\mu\text{Wcm}^{-2}\text{sr}^{-1}\text{nm}^{-1}$ @ 50 lines/sec Noise Equivalent Radiance at 635 nm |
| Dynamic Range | 12 bits |
| Line Rate | Up to 100 lines/sec |
| Recording | Digital helical scan tape recorder, 2 GByte or 2 hour capacity |
| Mass | 55 Kg |
| Power | 110 VAC, 2.5 A |
| Temperature | 5 to 40° C operating -20 to 60° C storage |

IMAGING SPECTROSCOPY AT THE FOCAL PLANE

Thomas S. Pagano, Carl F. Schueler,
and Loren M. Woody

Santa Barbara Research Center
Hughes Aircraft Company
75 Coromar Dr.
Goleta, CA 93117

Abstract

Novel, spectral separation technology lies at the heart of new instrument strategies being conceived for advanced earth observation. The simplification of the optical design with spectral separation on the focal plane instead of the optical system is demonstrated in a sensor design concept for low earth orbit of a high spectral and spatial resolution Environmental Imaging Spectrometer (EIS). The EIS would offer 126 spectral bands from 0.4 to 2.5 μm with bandwidths ranging from 6 to 40 nm. Excellent radiometric sensitivity and accuracy combined with the high spectral and spatial resolution of the EIS would enable researchers to analyze phenomenology of the atmosphere, ocean, and land.

Key Words: Low Earth Orbit, Imaging Spectroscopy, Mineralogical, Biological, "Red-edge", BRDF, EIS, Thematic Mapper, Sensor.

1. Sensor Requirements

Scientists have for many years expressed a desire to map the mineralogical and biological composition of the Earth's surface from space. The major impediment to this goal is access of sufficiently detailed spectral information to match spectral signatures of surface materials (Vane and Goetz, 1988). A major result has been the definition of the High-Resolution Imaging Spectrometer (HIRIS) instrument for Eos, which offers continuous spectral coverage at 10 nm resolution from 0.4 to 2.5 μm with a spatial resolution of 30 m from low earth orbit.

An examination of the reflectance spectra of rocks, soils, and vegetation (Figure 1) indicates that continuous spectral coverage at constant 10 nm resolution is not absolutely essential to perform the majority of the science objectives of imaging spectroscopy. Some of the spectral features of interest are sharp enough that 10 nm or even finer resolution is needed. However, sharp spectral features are not uniformly distributed throughout the entire 0.4 to 2.5 μm spectral range. As a result, it is possible to relax the spectral resolution requirement in portions of the spectral range without substantially compromising the spectral resolution requirement in portions of the spectral range without substantially compromising the spectral signature matching goal (Elvidge, 1987; Feldman and Taranik, 1988; Hutsinpillar, 1988). Table 1 delineates a set of bands that have been developed for both geological and biological feature identification

requirements developed by the Desert Research Institute.

The spectral features of green leaves and iron oxide minerals in the visible spectral region generally require 10 to 20 nm resolution, so that 15 nm bands appear adequate in this region. In the red, and near infrared region, however, shifts on the order of 5 nm in location of the sharply defined red edge of the chlorophyll pigment, as shown in Figure 1, offer information on the physiological condition of plants. Therefore, spectral resolution of 10 nm is not fully adequate here. The imaging spectrometer for this application should provide the finest resolution possible across the chlorophyll red edge region (0.66 to 0.80 μm). Further out in the infrared (0.80 to 2.5 μm), many features of interest can be discriminated with coarser spectral resolution, and these latter relaxed specifications reduce the cost and risk associated with sensor development. The 2.0 to 2.5 μm region contains a large number of fine spectral features from lignin, cellulose, and a variety of minerals; therefore, the imaging spectrometer for this region should have the finest spectral resolution and sampling interval possible.

A spatial resolution of 30 m for all channels and signal-to-noise ratio (SNR) of greater than 150:1 for surfaces having 50% albedo are required. Stereo and bidirectional reflectance distribution function (BRDF) measurement capabilities are required with a four-day minimum repeat coverage cycle. These requirements, provided by the Desert Research Institute, have been used by Santa Barbara Research Center (SBRC) to design an instrument concept called the Environmental Imaging Spectrometer (EIS).

2. EIS Instrument Concept

Figure 2 illustrates the EIS concept that fulfills the spectral requirements of Table 1, as well as the other spatial and radiometric requirements of the previous section. Figure 3 shows the EIS scene geometry. At the orbital altitude of 824 km, and spatial resolution of 30 m, an 11.52 km swath can be covered with an instrument field of view of 0.80 degrees and a cross-track detector array length of 384 elements. As a result, the EIS can use a pushbroom configuration without unduly taxing the optics field of view, focal plane size, data rates, or cooling requirements. The narrow swath and pushbroom configuration reduce sensor complexity significantly while still permitting data acquisition over any spatial region of interest by

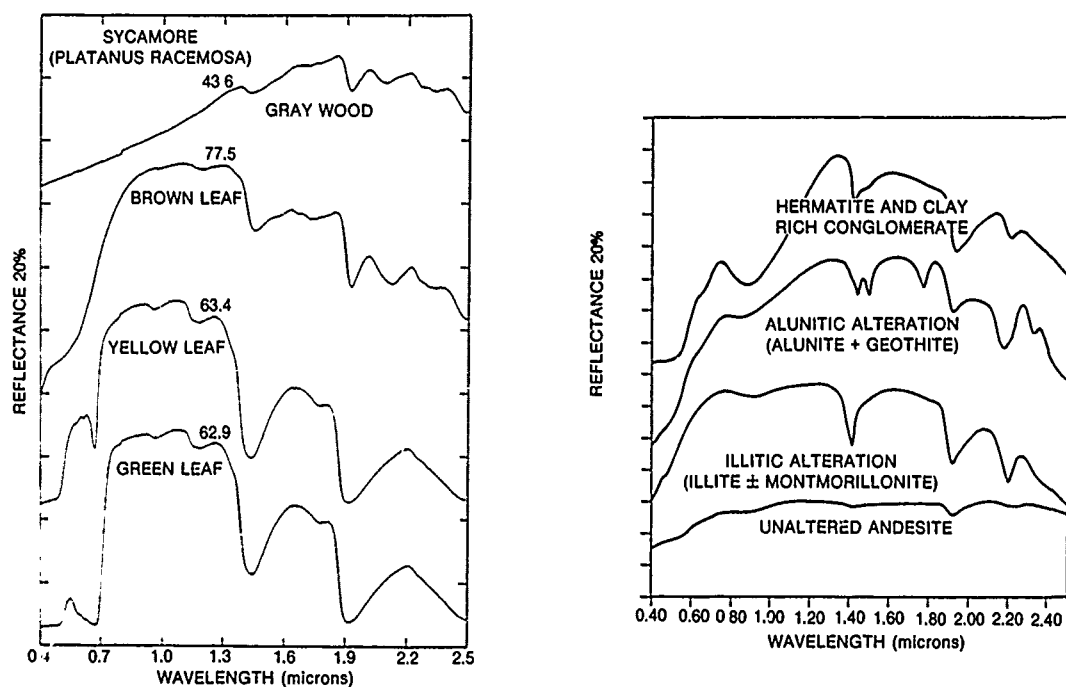


Figure 1. Reflectance spectra for sycamore (left) and for rock materials from the Virginia Range (Right) (Provided by Chris Elvidge of the Desert Research Institute.)

Table 1. ELS Performance Requirements

| Spectral Region | 1 | 2 | 3, 4 | 5 |
|------------------------------------|---------------------------|-------------------|--------------|--------------------------|
| Wavelength Range (μm) | 0.4-0.66 | 0.66-0.80 | 0.8-1.0-2.0 | 2.0-2.5 |
| Spectral Features of Interest | Leaves; "Red-Iron -Oxide" | "Red Edge" Shifts | Rocks; Soils | Lignocellulose; Minerals |
| Number of Bands | 18 | 25 | 42 | 41 |
| Spectral Sampling (nm) | 15 | 6 | 30 | 12.5 |
| Bandwidth | 10-14 | 6-8 | 20-40 | 20-25 |

using a two-axis pointing capability. The size of the instrument concept, 1.1 m (l) \times 0.9 m (w) \times 0.45 m (h) is smaller than would be an equivalent system that used a prism or grating spectrometer for spectral separation. A Thematic Mapper-type radiative cooler is used to cool the infrared detectors to 140K. The estimated weight is 90 kg and the total estimated power dissipation is 200 W. The raw instrument data rate at 100% duty cycle is estimated to be 109 Mbps for all channels.

A four-day repeat coverage can be attained by stepping the field of view in the cross-track direction to any position within ± 24 degrees. Along-track pointing to any position within ± 45 degrees is provided for measuring BRDF and for performing atmospheric correction. Pointing will also permit acquisition of stereo imagery. Pointing along track to an angle of 90 degrees is possible for viewing the on-board solar dif-

fuser. Pointing is achieved with a single elliptical two-axis pointing mirror, approximately 38 cm long in the cross-track direction and 22 cm wide in the along-track direction.

3. Focal Plane Spectral Separation

Special dielectric filter techniques, under development by Santa Barbara Research Center under IR&D, offer the flexibility and high performance of an imaging spectrometer in a very compact package. The specially designed filters are mounted directly on the detector array, avoiding the use of bulky and expensive grating or prism systems and their associated optics. A schematic comparison of the special SBRC filter system to a prism spectrometer system is shown in Figure 4. The filter system offers significantly reduced system design complexity. Although all bands are spatially imaged onto different ground locations

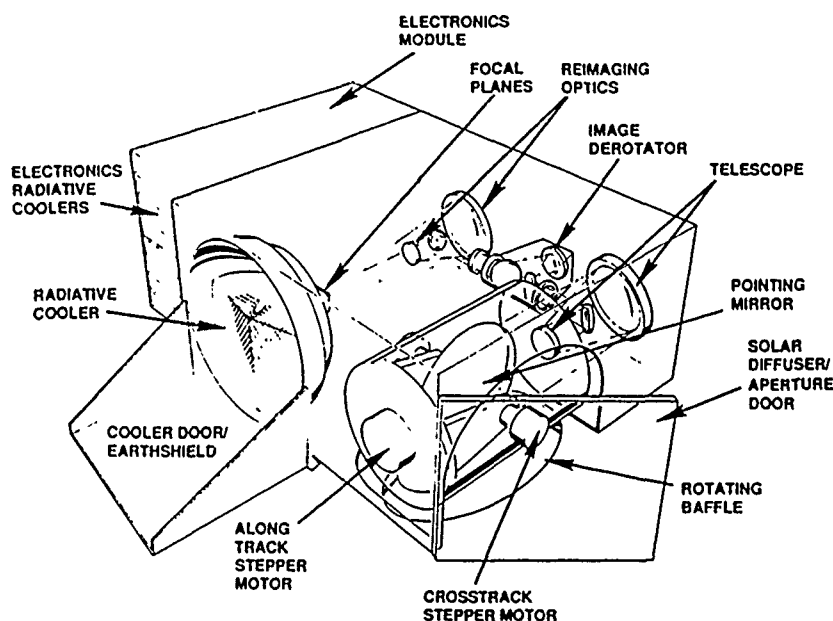


Figure 2. The EIS instrument is designed specifically to attain the environmental science goals.

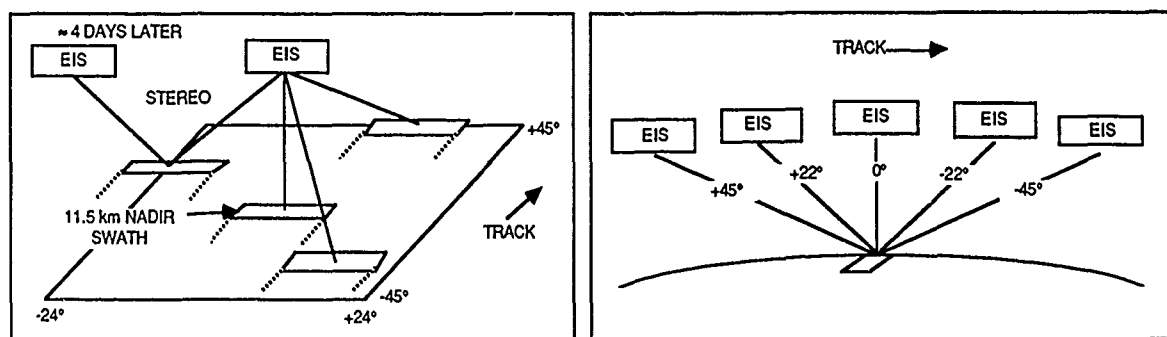


Figure 3. EIS two-axis pointing provides ± 24 degrees cross track and ± 45 degrees along track coverage.

as shown in the figure, their relative locations on the focal plane are fixed by the focal plane geometry; therefore, post processing taking account of the sensor ground track motion allows accurate band-to-band registration.

The sensor design requires two focal planes: (1) a buried-channel CCD imager (0.4 to 1.0 μm) and (2) a modular 388×80 HgCdTe imager (1.0 to 2.5 μm). Read-out in the CCD uses floating gate technology, while the HgCdTe focal plane uses SBRC-developed Multispectral Linear Array (MLA) technology. These high-performance focal planes result in excellent radiometric sensitivity. Five special SBRC dielectric filters, each custom-tailored to one spectral region in Table 1, are mounted to these focal planes. This configuration results in a total of 126 spectral bands from 0.4 to 2.5 μm . Figure 5 shows the resulting sampling interval and bandwidths achievable with this technology. The bandwidth variation shown in the figure is a result of the filter manufacturing technique. This variability of the bandwidth results in undersampling and over-sampling for a fixed sampling interval. This effect does not degrade the performance or calibration of the instrument.

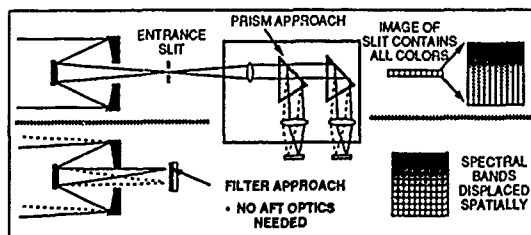


Figure 4. Schematic comparison of the filter approach to spectral separation to a prism spectrometer system.

4. Optics

To accomplish the EIS science mission, high image quality throughout the spectral range must be attained with high spectral transmittance. In the EIS optical design, a concentric, two-mirror Ritchey Chretien telescope was selected because it is compact and has excellent image quality throughout the field of view.

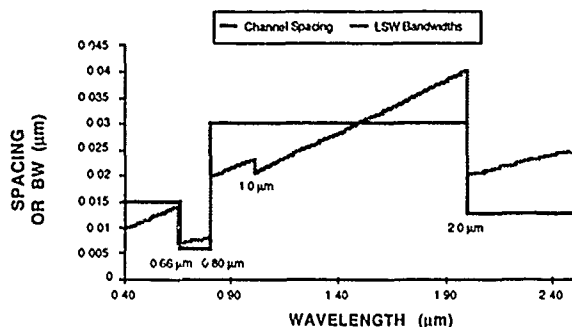
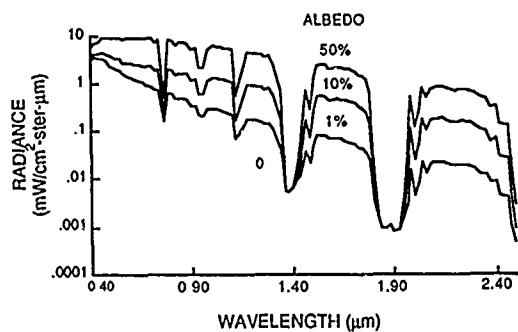
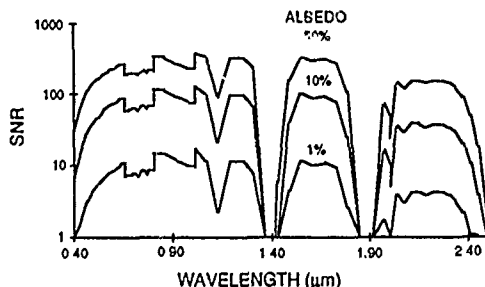


Figure 5. Sampling interval and bandwidth vs. wavelength for the EIS. The EIS has 126 spectral channels.

No prisms or gratings are used in the design since full spectral separation is achieved on the focal plane. An image derotator is incorporated to correct for the rotation that will be encountered in turning the pointing mirror in the along-track direction. A two-mirror reimaging system is employed to focus all bands within the dewar. A telecentric design uses two aspherics modified from a Ritchey Chretien form, providing an $f/4$ converging beam and a flat focal plane.



Radiance includes atmospheric contribution.



SNR after atmospheric correction, Plot includes every other channel

Figure 6. Predicted radiance at the sensor and SNRs, corrected for atmospheric radiance, with respect to wavelength at 1%, 10%, and 50% albedo.

5. Signal-to-Noise Ratio (SNR)

The predicted radiance levels at the sensor obtained

from LOWTRAN 6 and SNR after atmospheric correction are shown in Figure 6. Three radiance and SNR for different albedos; 1%, 10%, and 50% are plotted. The SNRs predicted for the EIS are achievable because:

- Along-track (pushbroom) scanning configuration results in long dwell times.
- There are broad spectral bandwidths in all regions, except for the 0.66 to 0.80 μm region, without compromising fine channel spacings.
- Focal planes have low detector dark current and readout noise.

6. Summary and Conclusions

The Environmental Imaging Spectrometer (EIS) concept uses state-of-the-art detector and spectral separation technology to provide high spatial and spectral resolution from low earth orbit in a compact instrument. The resulting concept has 126 spectral bands from 0.4 to 2.5 μm with varying resolution and sampling. Spatial resolution of 30 m is achieved in a 11.52 km swath. Pointing along-track and cross-track provides additional flexibility for acquiring stereo imagery, BRDF measurements, atmospheric correction, and rapid revisit. Designed to meet the science objectives of the user community while minimizing size, weight, power, data rate, and cost, the EIS is an attractive solution for high spatial and spectral resolution remote sensing from low earth orbit.

7. Acknowledgments

The authors would like to thank Dr. Christopher D. Elvidge of DRI for his assistance in defining the sensor requirements, Geoff Walter of SBRC for his continued efforts in developing the spectral filter technology, Richard Cline of SBRC for the optical design of the system, and Stillman Chase of SBRC for his editorial and technical comments.

8. References

1. Vane, G., and Goetz, A.F.H., "Terrestrial Imaging Spectroscopy", Remote Sensing of the Environment, Vol. 24, No. 1, 1988.
2. Elvidge, C.D., "Reflectance Characteristics of Dry Plant Materials", Proceedings of the Twenty-First International Symposium on Remote Sensing of Environment, ERIM, Ann Arbor, Michigan, October 26-30, 1987.
3. Feldman, S.C. and Taranik, J.V., "Comparison of Techniques for Discriminating Hydrothermal Alteration Minerals with Airborne Imaging Spectrometer Data", Remote Sensing of the Environment, Vol. 24, No. 1, 1988.
4. Hutsinpillar, A., "Discrimination of Hydrothermal Alteration Mineral Assemblages at Virginia City, Nevada, Using the Airborne Imaging Spectrometer", Remote Sensing of the Environment, Vol. 24, No. 1, 1988.
5. Goetz, A.F.H., and M. Herring, "The High Resolution Imaging Spectrometer (HIRIS) for Eos", IEEE Trans. Geoscience and Remote Sensing, Vol. 27, No. 2, 1989.

QUANTITATIVE DETERMINATION OF IMAGING SPECTROMETER SPECIFICATIONS BASED ON SPECTRAL MIXING MODELS

Alexander F. H. Goetz
Joë W. Boardman

Center for the Study of Earth from Space (CSSES)
CIRES, University of Colorado, Boulder, CO, USA 80309

ABSTRACT

New sensor development requires a translation of scientific requirements into functional requirements to guide the engineering design of an instrument. We have developed a model to aid in the design of the High Resolution Imaging Spectrometer (HIRIS), a facility instrument slated for the EOS mission. The model is based on the concept that scientific requirements for an imaging spectrometer can be stated in terms of the absolute error allowable in determining the abundance of a given surface material in the presence of others. Pure end member materials are represented by a library of spectral reflectance curves. For any library and a required abundance accuracy, the model yields values of spectral and radiometric resolution required. The latter can be expressed in terms of signal-to-noise ratio. These parameters can be translated into optics performance and data rates that ultimately determine the cost of the instrument.

Key Words: imaging spectrometry, HIRIS, spectral unmixing, instrument design, EOS

INTRODUCTION

The challenge of earth remote sensing from space and also its curse is that the results are of interest to such a wide variety of users and disciplines. Since the advent of Landsat in 1972, this phenomenon has led to the "design by committee" technique for each new imaging sensor. The discipline scientists in these committees are typically asked to state their requirements for spatial resolution, swath width and number and placement of spectral bands. More often than not, these requirements, or more aptly desires, were comprised by engineering realities, limitations imposed by weight, power, data rate, pointing stability and technology. In the last decade, advances in technology have greatly increased the number of engineering options available to the discipline scientist and made the choices concomitantly more difficult. "Wer die Wahl hat, hat die Qual" aptly expresses the dilemma faced by committee members confronted with decisions on the new breed of sensors, imaging spectrometers.

Imaging spectrometry of the earth (Goetz, et al, 1985) has greatly enhanced our ability to identify surface cover directly, through analysis of complete reflectance spectra on a pixel basis (Vane and Goetz, 1988). The acquisition of reflectance data in contiguous narrow spectral bands has removed the requirement to select a limited number of bands in the sensor and has shifted the responsibility to the investigator during data analysis. In spite of advances in technology, data rates are still a limiting factor in data acquisition from orbit. Therefore, trade-offs are necessary among spatial and spectral resolution, swath

width, and encoding. Signal-to-noise ratio is another cost driver in imaging spectrometer design that must be addressed in an objective manner.

The development of an instrument as complex and costly as the High Resolution Imaging Spectrometer (HIRIS) for the NASA Earth Observing System (EOS) (Goetz and Herring, 1989, Goetz et al, 1987) requires a quantitative approach to the development of scientific requirements and the related functional requirements (Goetz and Calvin, 1987). In this paper we describe a model that relates scientific requirements stated in terms of the absolute error allowable in determining the abundance of a given surface material in a pixel, in the presence of others, to instrument requirements for spectral and radiometric resolution. The technique used is spectral unmixing of a pixel composite spectrum from a library of spectral reflectance curves of pure end-member materials.

HIRIS

The High Resolution Imaging Spectrometer is designed to acquire images in 192 spectral bands simultaneously in the 0.4-2.5 μ m wavelength region. HIRIS is a targeting rather than a continuous acquisition instrument and obtains high spatial and spectral resolution images in a 24 km swath with a 30m GIFOV in vertical viewing. 2-Axis pointing is proposed which will allow image acquisition at -30° to $+52^{\circ}$ down-track and $\pm 26^{\circ}$ cross-track. The raw data rate of the instrument is 405Mbs. The high spectral resolution will make it possible to directly identify surficial materials such as rocks and soils and suspended matter in water, and HIRIS opens up the possibility of studying biogeochemical processes in vegetation canopies. HIRIS will be used in conjunction with the Moderate Resolution Imaging Spectrometer (MODIS) as a multi-stage sampling system.

The 405Mbs instrument data rate exceeds the capability of the EOS platform data system and various options are being studied to reduce the rate. All options affect the scientific requirements, and this is one area of the design that requires an objective approach to defining the relationship between scientific and engineering requirements. An additional benefit is more harmonious relationships among team members. The focus in the model under development is the interaction between the requirements for spectral and radiometric resolution also called signal-to-noise ratio. These parameters affect the data rate as well as the overall size and cost of the instrument.

APPROACH

Spectral unmixing models were used to quantitatively address the effects of typical end-user scientific requirements on instrument design parameters. The individual and combined effects on the accuracy of spectral unmixing due to changing

signal-to-noise ratio, spectral resolution and unmixing library degeneracy were examined using a suite of synthetic data sets. A spectrally mixed data set was created for each combination of the three variables. The four noise levels, five spectral resolutions and the four different mixing libraries used resulted in eighty unique spectrally mixed simulated imaging spectrometer data sets. These spectrally mixed synthetic data were unmixed through inversion of a linear mixing model. To ascertain the effects on the unmixing accuracy caused by the changing parameters the unmixing results were compared to the true end-member abundance patterns used in creating the synthetic data. In this manner the unmixing errors can be plotted as a function of the three variables: signal to noise ratio; spectral resolution; and library degeneracy.

The key elements used in the creation of the simulated data sets were the six spatial abundance test patterns and the noise free, ideally sampled mixing end-member spectra. The six, 200 sample by 200 line, spatial abundance test patterns are simple geometric shapes such as cones, wedges and step functions. These were used to determine, on a pixel by pixel basis, the abundance of each end-member to combine to form that mixed spectrum. The six spectra to be associated with the six spatial abundance patterns are synthetic reflectance spectra covering a wavelength region of fifty arbitrary units. Each has a single gaussian absorption feature with zero reflectance at the feature center and a full width at half maximum (FWHM) of ten wavelength units (Figure 1). Thus, through a convolution of the six chosen mixing spectra with the six spatial abundance patterns, a simulated reflectance data set is created with dimensions of 200 samples by 200 lines by the number of channels used to sample the wavelength range of the spectra.

In the course of creating the suite of synthetic data sets, the signal to noise, spectral resolution and library degeneracy was varied. A random noise spectrum with a gaussian distribution was added to each mixed spectrum. The signal-to-noise is defined here as one hundred percent reflectance divided by the standard deviation of the gaussian noise spectrum. Four different standard deviations were used, 0., 0.002, 0.02 and 0.2. According to our definition these correspond to signal-to-noise ratios of infinite, 1, 500:1, 50:1 and 5:1. These span the range of such values for actual instruments. Changing the spectral resolution was accomplished by convolving each mixing end-member with a triangular filter. These band-pass filters are separated by a spectral distance equal to one half their full width. This corresponds to spectral band-passes that intersect at their half maxima resulting in a spectral sampling interval which is one half the spectral resolution, also known as critical sampling. Five different critical sampling intervals were used here, 1, 2, 3, 4 and 5 wavelength units. To model the effects of increasing library degeneracy, decreasing spectral separability, the center of the absorption feature of the first end-member was moved closer to that of the second. Four different feature separations were used. Initially all the features were separated by ten wavelength units. The distance between the feature centers of the first and second end-members was decreased to 5, 2.5 then finally 0.5 wavelength units. By carrying out all of the possible permutations the eighty different cases were synthesized.

Once the mixed data sets were synthesized, unmixing using the proper noise-free end-members was accomplished by using an unconstrained, linear, least squares inversion technique. Singular value decomposition was used to invert the library matrix (Boardman, 1989). Singular value decomposition decomposes the library matrix into the product of two column orthogonal matrices of singular vectors and a square diagonal matrix of singular values. Once orthogonalized the inversion of the library matrix is straightforward. This technique was used for the matrix inversion since it orthogonalizes the end-members, allows insight into the

spectral degeneracy of the library and could be used to balance resolution and solution variance during the inversion process. The synthesized spectra were multiplied by this inverted library matrix. This unmixing results in an estimated abundance for each of the six end-members for every pixel. When all of the results were spatially reassembled a set of estimated abundance spatial distribution patterns were synthesized. These could be directly compared to the true abundance spatial patterns used to do the mixing. A total unmixing error for each end-member was accumulated as the square root of the mean of the squared differences between the estimated and true abundances (RMS error). In this manner a set of six unmixing errors, one for each end-member, was produced for each of the eighty cases.

RESULTS

The results for the inversion of the synthetic data sets formed by mixing the idealized artificial spectra indicate a linear relationship between unmixing error and both the noise level in the data (inverse of signal-to-noise ratio) and the spectral resolution of the instrument. The slopes of these linear functions are determined by the separability of the end-members in the unmixing library. For a fixed noise level, libraries with a greater degree of spectral separability are less sensitive to a given decrease in spectral resolution than libraries that are more spectrally degenerate (Figure 2). A plot of this relationship for all four libraries used, at a fixed noise level, shows the unmixing error as a function of both feature separation, a measure of separability of the synthetic end-members and spectral sampling interval (Figure 3). Error values for higher or lower noise levels can be found by a simple linear scaling of the values for a known noise level, since the error versus noise sensitivity has a zero intercept at a zero noise level.

The procedures outlined here are directly applicable to studies involving real mineral spectra. An example of an application to real mineral spectra is shown in Figures 4 and 45. The signal to noise level is fixed at 50:1 and the sensitivity to spectral sampling interval is examined. Here six different minerals are chosen for the mixing library. In the first library all of the minerals are relatively spectrally separable. In the second library the substitution of dolomite for muscovite increases the degeneracy of the library since its spectrum is very similar to calcite. Note the increased sensitivity to degradation of spectral resolution when using this less separable library. Using the same two libraries, with fixed spectral resolution, the sensitivity to increased noise is shown in Figures 6 and 7. The increased degeneracy of the second spectral library, due to the inclusion of dolomite, increases the sensitivity of the system to noise. These examples point out the importance of judiciously choosing the library of interest.

CONCLUSIONS

The results of this study can be applied directly to both instrument design considerations and feasibility studies for planned imaging spectrometer applications. For a given library, surfaces of equal unmixing error can be determined as functions of instrument noise level and spectral resolution. This family of parameter combinations represents instruments that should all have the same unmixing accuracy for that library. By applying a cost function to each parameter the optimum combination of low cost and high unmixing accuracy can be determined. For an end user, limited to an existing instrument with fixed parameters, studies of this sort using the unmixing library of interest should allow estimation of feasibility and absolute unmixing accuracy for each end-member. For an end user of an instrument with variable parameters, such modeling can allow optimization of the instrument configuration for a given application. In such an application the end user would pick an unmixing library and

determine the noise level in the data. Then, after resampling the library end-members to match the spectral resolution of the instrument, best case unmixing error estimates could be made for each end-member.

The errors examined in this unmixing model represent the best case scenario for any such application. They illustrate the upper limit of unmixing accuracy that could be achieved for a given library and instrument combination. In real applications other sources of error will certainly reduce the true achievable unmixing accuracy. Such other sources of error include non-ideal libraries (those that are not all inclusive or those that include spurious members,) and errors in calibration of the data and in the conversion to reflectance. Uncompensated atmospheric and topographic effects also will decrease the unmixing accuracy. However, we believe that our results are useful in both instrument design considerations and end user analysis and interpretation since they allow a quantitative assessment of the effects of instrument parameters on a typical application of imaging spectrometry data.

ACKNOWLEDGEMENTS

This study was funded in part by NASA/JPL contract # 958039.

CITED REFERENCES

1. Boardman, J.W., "Inversion of imaging spectrometry data using singular value decomposition", IGARSS proc. 1989, Vancouver, this publication, 1989.
2. Goetz, A.F.H. and Calvin, W.M., "Imaging spectrometry : spectral resolution and analytical identification of spectral features", Imaging Spectrometry II, Gregg Vane, Editor, Proceedings SPIE, v. 834, p. 158-165, 1987.
3. Goetz, A.F.H. et al, "High Resolution Imaging Spectrometer: Science Opportunities for the 1990's", NASA Earth Observing System, v. 2c, 1987.
4. Goetz, A.F.H. and Herring, M., "The High Resolution Imaging Spectrometer (HIRIS) for EOS", IEEE Transactions on Geosci. and Rem. Sens., v. 27, p. 136-144, 1989.
5. Goetz, A.F.H., Vane, G., Solomon, J.E., and Rock, B.N., "Imaging spectrometry for earth remote sensing", *Science*, v. 228, p. 1147-1153, 1985.
6. Vane, G. and Goetz, A.F.H., "Terrestrial imaging spectroscopy", *Remote Sensing of the Environment*, v. 24, p. 1-29, 1988.

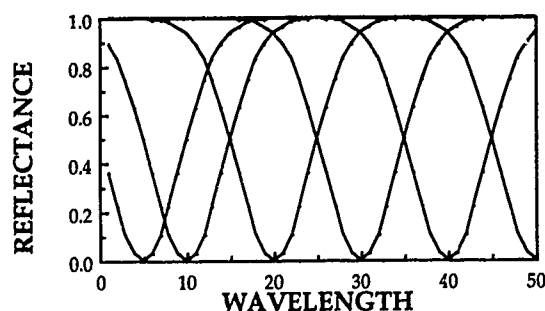


Figure 1.

This is an example synthetic spectral mixing library used in this study. In this library endmembers #1 and #2 are only separated by 5 wavelength units, modeling decreased spectral separability.

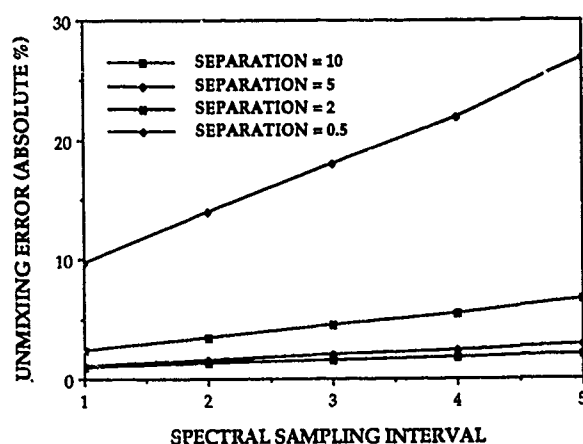


Figure 2. Unmixing error as a function of spectral sampling interval using artificial end-members and a fixed S/N of 50.

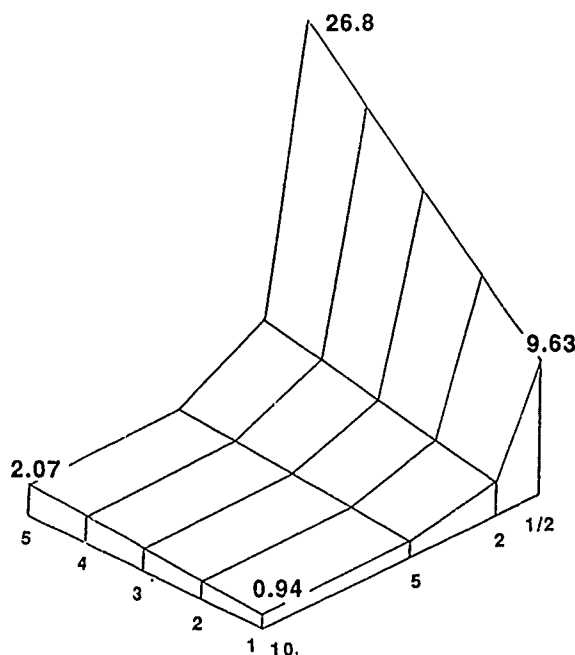


Figure 3. Unmixing error as a function of spectral sampling interval and feature separation between end-members 1 and 2, S/N fixed at 50.

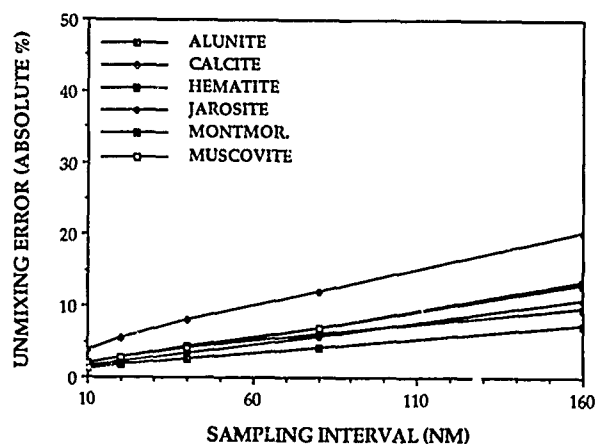


Figure 4. Unmixing error for library #1 as a function of sampling interval with S/N fixed at 50.

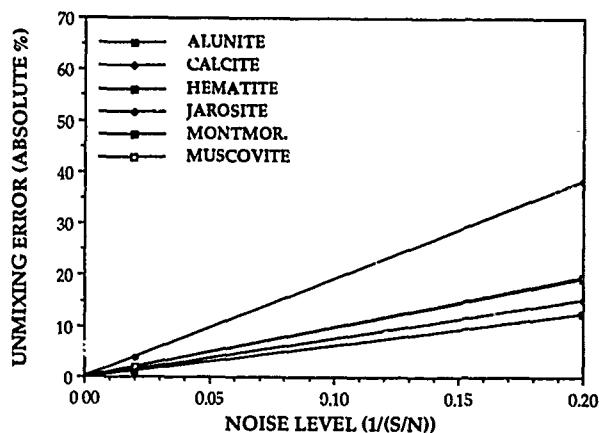


Figure 6. Unmixing error for library #1 as a function of noise level, spectral sampling interval is fixed at 10 nm.

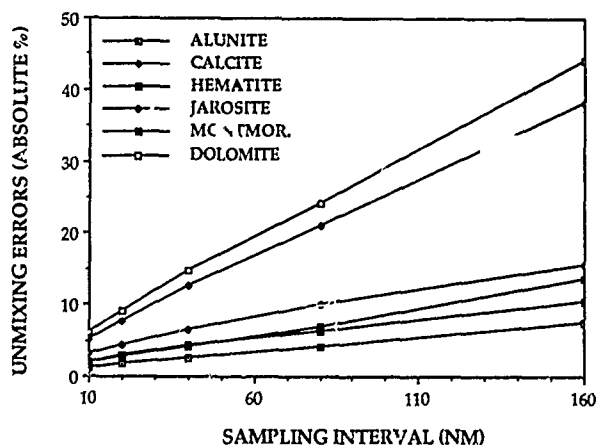


Figure 5. Unmixing error for library #2, same as library #1 except dolomite has been substituted for muscovite, S/N fixed at 50. Note the higher sensitivity to spectral resolution for this more degenerate library. Calcite is spectrally similar to dolomite.

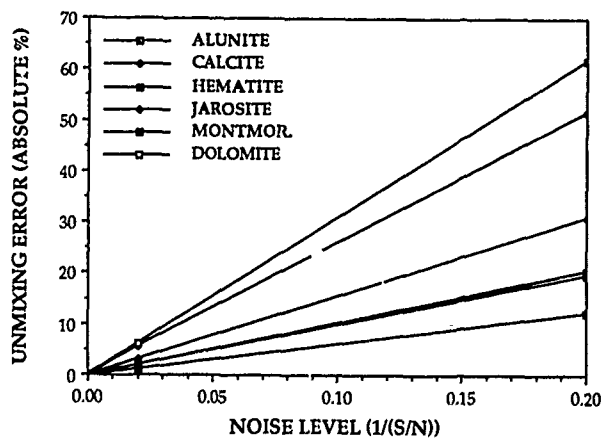


Figure 7. Unmixing error for library #2, same as library #1 except dolomite has been substituted for muscovite, spectral resolution fixed at 10 nm. Note the higher sensitivity to noise for this more degenerate library. Calcite is spectrally similar to dolomite.

FLUORESCENCE LIDAR DESIGN: OPTIMIZATION CRITERIA

Giovanna Cecchi, Luca Pantani, Marco Romoli

Istituto di Ricerca sulle Onde Elettromagnetiche - Consiglio Nazionale delle Ricerche
Via Panciatichi, 64 - I50127 FIRENZE (Italy)

ABSTRACT

This paper describes the optimization criteria for the design of fluorescence lidars, in order to exploit their full potential in operative environmental remote sensing; the optimization of both the whole system and its separated parts was taken into account. The criteria are the result of a 8 year research program carried out at IROE - CNR in Firenze, Italy, by means of theoretical, laboratory, and field investigations. The fluorescence lidar systems which were built at the IROE are described at the end of the paper.

1. INTRODUCTION

Fluorescence lidars have wide application fields in environmental remote sensing, particularly in the detection of sea parameters like oil pollution, phytoplankton, suspended and dissolved substances (Measures, 1983), and vegetation parameters like photosynthetic activity and stresses (Cecchi et Al., 1985).

The fluorescence Lidar technique is still in development and various systems were built during the past years, sometimes with very different structures and performances (Measures, 1983).

From an operative point of view, a compact, light weight, and low power consumption system is needed, which can be installed on different platforms, like trucks, aircraft, helicopters, and ships. A cost-benefit analysis of the different techniques, such as high spectral resolution detection, time decay measurements, and simultaneous detection of passive and active spectra, has also to be taken into account for the different applications. In the following paragraphs the optimization criteria adopted by IROE for the design of its system are discussed.

2. SYSTEM OPTIMIZATION

The information on the target is carried out by both the spectral and temporal behavior of the backscattered signal. The detection of the fluorescence time decay from moving platforms requires sophisticated techniques, such as the use of subnanosecond laser pulses with consequent fast detection systems. The information contained in the fluorescence time decay is useful in discriminating the composition of oil films floating on water surface (Measures et Al. 1974). However the variations due to sea waves in the distance between the lidar and the film could introduce errors in the absolute measurement of the time decay. In any case a

discrimination between the oil types can be also achieved by the analysis of the fluorescence spectrum with correlation (O'Neal et Al., 1981) or differential fluorescence techniques (Cecchi et Al., 1986). For what concern the monitoring of vegetation and phytoplankton, the usefulness of the time decay measurement is still under analysis; anyway the thickness of the target will introduce an ambiguity which is of the same order or higher than the time decay itself. In fact, by considering a chlorophyll fluorescence time decay in the range of few nanoseconds (typically 2-3 ns), the corresponding spatial range is some tenth of centimeters (less than 50 cm). So, in our opinion, for both applications (sea and vegetation) the measurement of the fluorescence time decay involves a hard complication of the system, while the same kind of information can be extracted directly from the fluorescence spectrum. As a consequence this technique was abandoned in the design of IROE fluorescence Lidars.

In the past two main design philosophies were adopted for the detection of the spectral characteristics of the target:

- a - few receiving channels with separate interference filters and photodetectors,
- b - many receiving channels with a grating spectrometer and an array detector.

The first solution gives information only on a small number of fixed fluorescence bands, but allows an easy detection of depth resolved signals. The second one gives information on the whole fluorescence spectrum with high resolution; moreover it allows the detection of laser excited spectra and passive reflectance spectra of the same target. This feature is very attractive, because the reflectance spectrum can supply relevant complementary information on the target, like the leaf area index (LAI). On the other hand the detection of depth resolved signals is more complex.

Since the goal of the research carried out at IROE was the development of fluorescence lidar which could perform the detection of the largest possible set of environmental parameters, solution (b) was adopted in the design.

The concept of a main computer which controls via software the whole system, operates signal preprocessing, gives a real time data display, archives the data on a suitable support, and eventually processes the data, was selected for the system management. This solution offers the maximum of adaptability of the system and is the easiest one for the operator, particularly in narrow surroundings like

inside a small plane.

2.1. SOURCE

The selection of excitation wavelengths, and consequently of the optimal laser sources, was done on the basis of fluorescence efficiency and absorption spectrum of the target, absorption coefficients of propagation media, and laser characteristics.

The propagation medium is the atmosphere, while water is involved when the measurement of water column parameters is performed. The strong absorption of the atmospheric ozone in the ultraviolet (UV) restricts the choice to excitation wavelengths higher than ≈ 280 nm.

Various laboratory results (Measures, 1983) have already shown that the most suitable laser wavelength for the remote sensing of thin oil films has to be chosen in the 300-350 nm range; while a wavelength in the blue region of the spectrum (≈ 480 nm) is needed for water quality measurements (i.e. water turbidity, gelbstoff, etc.). A 480 nm wavelength is also the best excitation for chlorophyll *a* remote sensing, either in terrestrial vegetation, or algae, or phytoplankton. A possible source is an UV laser in the abovementioned range and a dye laser working at 480 nm, pumped by the UV laser itself. The dye laser can be an untuned one because of the wide absorption bands of the species which have to be detected.

The UV laser sources which are available are: triplicated Nd:YAG, Nitrogen, XeCl excimer. The N_2 laser was adopted in early fluorescence Lidars, then the other two sources were preferred, because of the low energy per pulse of N_2 laser. The IROE Lidar system adopted a XeCl excimer laser, which lases directly in the UV (308 nm).

2.2. OPTICS

The Lidar optical part is composed of the receiving optics and the dispersive system. A basic point is the dispersive system, which can be a set of beam-splitters with interference filters or a spectrometer, depending on the kind of detection system adopted. As already previously discussed, IROE Lidar uses a spectrometer dispersive system.

The receiving optics is usually a reflection telescope, whose characteristics have to match those of the laser and detection systems. In particular the telescope numerical aperture (NA) has to match that of the spectrometer, and its Field Of View (FOV) has to match the divergency of the laser beam. If the Lidar detects also passive reflectance spectra, the matching between the FOV and the laser beam divergency is compulsory in order to assure that the reflectance spectrum corresponds only to the area, which is illuminated by the laser.

A spectrometer hardly matches the collected spot-size, because of the shape of its narrow entrance slit. Further, the dispersed light has to match the detector size. Typically, the slit is hundreds μ m wide and ≈ 2 cm high, while the detector array element is 25 μ m wide and 2.5 mm high; so the detected spot size turns to be WdH , where W is the width of the entrance slit and H is the height of the detector array (i.e. $\approx 100 \mu$ m \times 2.5 mm). The ideal situation should be a focal spot, having the same dimensions. This condition is hardly achieved, even if the rectangular shape of XeCl excimer laser turns useful, when this excitation wavelength is applied.

Generally, the collected radiation has a circular shape; two cases are possible: or the image diameter is equal to the slit-width, or the image diameter is

equal to the array height. In both cases the signal to noise ratio is non-optimal, in fact in the first one, only a small part of the detector is irradiated, in the second one most of the energy is missed at the slit.

A way of overcoming this problem is the use of an optical interface between the telescope and the entrance slit, which transforms the circular shape into a rectangular one, keeping the area constant. An optical fiber cable can do this job. On one side the cable end has a circular shape for matching the telescope focal spot, while the other end has a rectangular shape, for matching the spectrometer slit and the array height. The signal to noise ratio improvement, due to the optimal coupling between the telescope and the entrance slit, largely compensates the insertion losses.

This solution allows an easy alignment of the system and, which is of relevance, permits a complete electrical insulation between the laser source and the detection system.

As a conclusion, a high resolution fluorescence Lidar could be realized accordingly to the following items:

- Both active and passive spectra acquisition.
- Data acquisition and Lidar operation computer controlled.
- Compact, low weight UV+VIS laser system
- Optical interface between the receiving optics and the detection system
- Grating spectrometer + array detector

3. IROE FLIDAR(*) PROJECT

On the basis of the previous considerations, a conceptual sketch of a fluorescence Lidar, which utilizes a Xe-Cl excimer laser and an excimer pumped dye laser as source, a properly designed receiving optics, a spectrometer, and an intensified gatatable photodetector array as receiver, was chosen as the best solution.

An actual realization of this system is the FLIDAR 2 (Castagnoli et Al., 1986), operating at our Institute (Fig. 1). The new FLIDAR 3 is under development and will be ready by the end of this year.

The laser sources of FLIDAR systems were expressly designed (Castagnoli et Al., 1986): in fact commercial excimer lasers are too bulky for operating on moving platforms. The excimer laser is battery powered and the overall dimensions are 80 cm \times 40 cm \times 35 cm (high voltage supply included); the weight is 60 Kg. (batteries included). Also the dye laser was expressly designed, it is an untuned one with overall dimensions of 30 cm \times 20 cm \times 12 cm.

The FLIDAR 3 is conceptually similar to the FLIDAR 2 and uses the same lasers, spectrometer, and detection system, while the receiving optics is now under design (Cecchi, 1989). Both laser sources are mounted on the same rigid plane with a mirror-selection of the laser wavelength. The laser system and the receiving optics are mounted on a platform which can be installed on an azimuth/elevation scanning fork for ground and truck operations, or directly inside an aircraft or a ship. The spectrometer, the detection system, and the main computer are hosted in a separate rack, only fiber optics cables connect the rack to the laser/receiver block and the two blocks are separately battery powered. This solution allows at the same time an easy alignment between the laser and the receiver, an easy alignment of the optical receiver with the detection system, and a very high EMI protection.

(*) FLIDAR is a registered trade-mark of CNR



Fig. 1 - FLIDAR - 2 in operation on the oceanographic ship "MINERVA" during the MARET campaign (October 1987)

R.C.Harney Editors), SPIE Vol.663, SPIE, Bellingham, pp212-216, 1986

2. Cecchi G., L.Pantani, I.Pippi, R.Magli, P.Mazzeinghi: "Vegetation remote sensing: a new field for Lidar applications", ECOOSA '84, SPIE Vol.492, pp180-185, SPIE, Bellingham Wa, 1985.
3. Cecchi G., L.Pantani, P.Mazzeinghi, A.Barbaro: "Fluorescence Lidar Remote Sensing of the Environment: Laboratory Experiments for the Characterization of Oil Spills and Vegetation", in Optoelectronics in Engineering (W.Waidelich Editor), pp652-655, Springer Verlag, 1986.
4. Cecchi G., P.Mazzeinghi, L.Pantani, M.Romoli: "Optimization of the receiving optics of a fluorescence Lidar", preparation, 1989.
5. Measures R.M., H.R.Houston, D.G.Stephenson: "Laser Induced Fluorescence Decay Spectra. A New Form of Environmental Signature", Optical Engineering, 13, pp494-454, 1974.
6. Measures R.M.: "Laser Remote Sensing", John Wiley & Sons, New York, 1983.
7. O'Neal R.A., F.E.Hoge, M.P.F.Bristow: "The Current Status of Airborne Laser Fluoresensing", 15th International Symposium on Remote Sensing of the Environment, pp379-389, Ann Arbor, Michigan, 1981.

REFERENCES

1. Castagnoli F., G.Cecchi, L.Pantani, I.Pippi, B.Radicati, P.Mazzeinghi: "Fluorescence Lidar for land and sea remote sensing", in: Laser Radar Technology and Applications (J.M.Cruickshank &

MULTI-FIELD-OF-VIEW LIDAR FOR SINGLE ENDED MEASUREMENT OF EXTINCTION COEFFICIENTS

P.E. LaRocque and A. Ulitsky

Optech Incorporated
701 Petrolia Road,
Downsview, Ontario
Canada

A compact lidar system has been built to acquire backscattered Nd:Glass laser returns from ranges up to 10km. Four receiver fields of view are obtained simultaneously: 5, 25, 50 and 75 mrad. The four channels are logarithmic amplified (a dynamic range of 5 orders of magnitude), digitized, stored and displayed by a personal computer. Extensive computer control has been designed into the system. A marine radar has been incorporated which automatically prevents the laser from firing when an object is detected in the non-eyesafe zone. The movable transceiver provides 90° rotation in two axes, stands only 1.4 meters high and can pass through doorways.

Keywords: Lidar, Extinction Coefficients, Multiscattering

Lidar systems have been used for many years to determine extinction coefficients of the atmosphere. However, the usual solution method requires a relation between the backscatter and extinction coefficients and the specification of a boundary value of the extinction coefficient, preferably at the far end of the lidar range. These parameters are generally not derivable from the backscattered signal and are usually obtained from a model imposed upon the atmospheric conditions. This technique can result in large errors and obviously is not suitable for general weather conditions.

A new method has been proposed (Bissonnette, 1986) based on a multiscattering approach. The lidar returns are recorded simultaneously at different fields of view and by ratioing these returns the backscattering coefficient can be factored out. Also, the slope of the logarithm of the ratios is proportional to the aerosol forward-scattering coefficient. With two ratios from three fields of view the forward scattering coefficient can be determined. This coefficient is directly related to the extinction coefficient by the albedo, within a few percent. Thus the extinction coefficient can be derived from the lidar data without the need for boundary values or the backscatter to extinction profile. In this paper we will describe a compact lidar system developed to confirm this multiscattering method.

The unique feature of this lidar is the simultaneous acquisition of the backscattered signal at four different fields of view: 5, 25, 50 and 75 mrad. A six inch diameter, f/1.33 telescope collects the backscattered light and focusses it onto a silicon PIN diode comprising four concentric active areas. The four channels are logarithmic-amplified, digitized by four 32 MHz waveform recorders and stored in a

personal computer. The lidar transmitter is a Nd glass laser with up to 2 joules of output energy. The transmitter and receiver are located on a moveable transceiver pedestal providing 95° rotation in elevation and $\pm 45^\circ$ rotation in azimuth. The transceiver stands only 1.4 meters high and can pass through standard doorways. It is separated from auxiliary equipment by seven meter long cables. As shown in Figure 1 this equipment consists of two 19" racks each under one meter in height plus a digitizing unit and a personal computer.

Extensive computer control has been designed into the system. Scanner control, system alignment, system calibration and lidar data acquisition are among the menus available on the computer terminal.

The scanner can be positioned in one of two ways by the computer: by specifying the angles desired or by moving a cursor within a grid. If the operator wants to align the system on a particular target, a hand held manual remote control is available which can be used in conjunction with a sighting scope.

Coaxial alignment of the transmitter and receiver is very important for system performance. With the use of a HeNe laser, the operator can call up the alignment menu and follow the simple steps which determine the optical axis of the receiver telescope to better than 0.2 mrad. The transmitter can then be aligned to the receiver optical axis.

Calibration of the system is achieved by placing the cover on the telescope and connecting two cables. A module on the telescope cover contains a light emitting diode and driver circuitry which illuminates the detector with uniform intensity. Interchannel calibration of the detector alone or of the detector with log amps can be completed within minutes by following the software menus. These calibration

constants are then saved and automatically stored with subsequent lidar data.

Acquisition of lidar data is prefaced by a run parameter menu. The operator selects the acquisition time (up to 2K samples per channel), an acquisition delay if desired and the number of laser pulses to average. Once the run parameters are entered, the acquisition key is depressed. The laser shutter will open and remain open until the specified number of laser pulses is emitted. The laser energy and detector temperature are reported for every laser fire. The shutter will then close and the data will be displayed immediately and stored along with the date, time, temperature, laser energy and relevant run parameters.

The high output energy of the laser results in the lidar not being eyesafe for a distance of several kilometers. In order to ensure safe operation a marine radar was incorporated into the system to automatically prevent the laser from firing when an object is detected within this non-eyesafe zone. There are also 'close laser shutter' controls at both the transceiver and equipment rack stations.

Details of the system design will be discussed and initial lidar returns will be shown. Figure 2 is representative of lidar data and Table 1 summarizes the system specifications. It is also hoped to present analysis of the multiscattering results obtained from initial field measurements with the system.

References

1. Bissonnette, L.R., "Multiple-Scattering Laser Propagation Model and Comparison with Laboratory Measurements", DREV R-4422/86 and "Multiscattering Lidar Method for Determining Optical Parameters of Aerosols", DREV R-4430/86.

Authors would like to acknowledge funding by Defence Research Establishment Valcartier (DREV), DND.

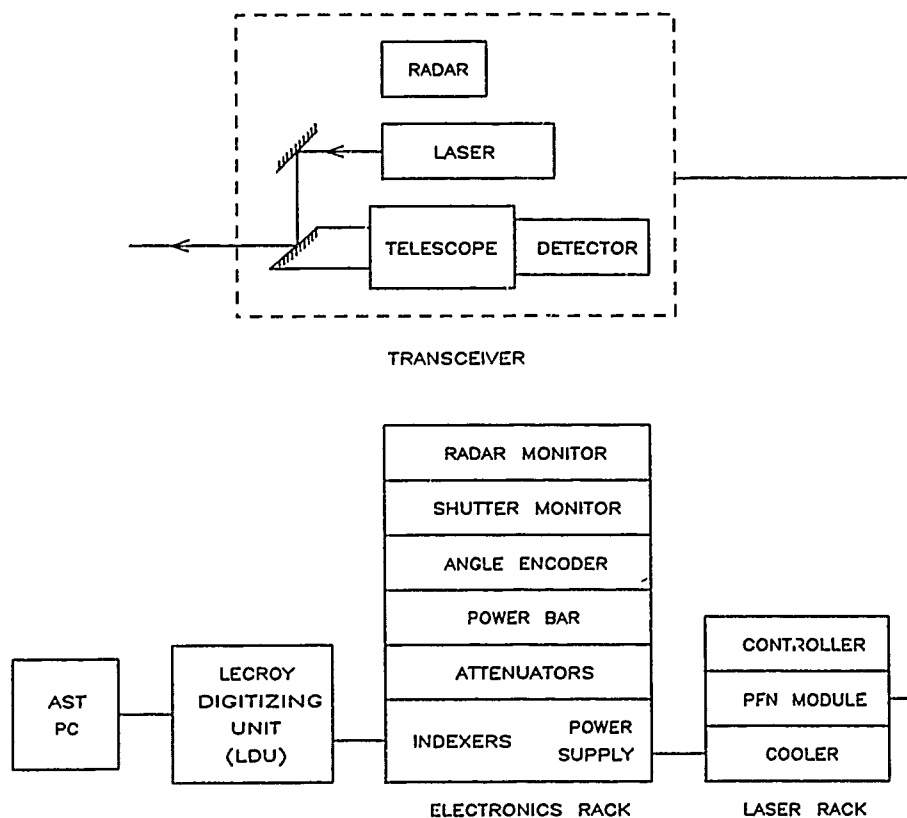


Figure 1. MFOV LIDAR SCHEMATIC

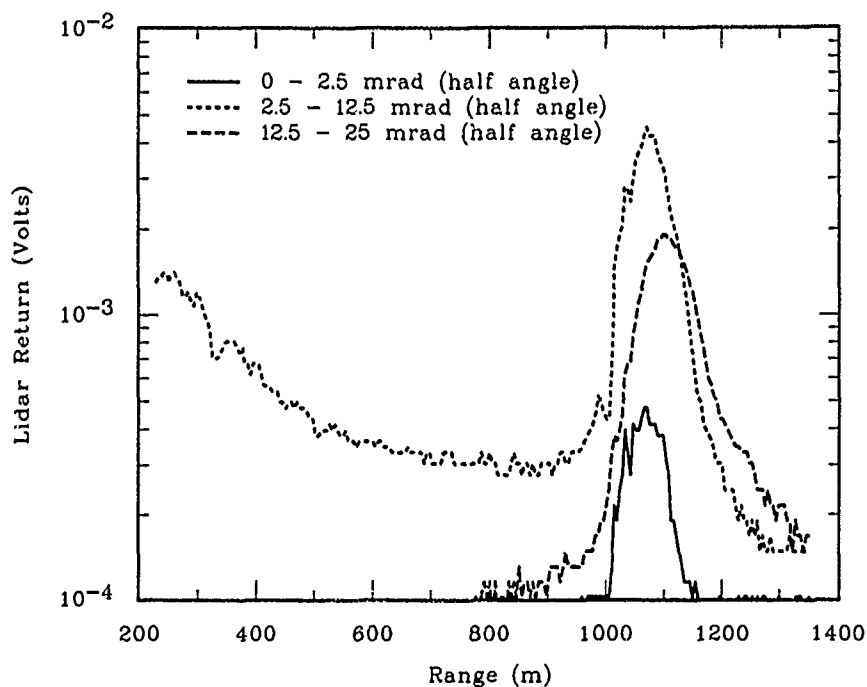


Figure 2. MFOV LIDAR DATA

Table 1
Multi-Field-of-View Lidar Specifications

Transmitter

| | |
|-----------------|------------|
| Laser Type | Nd:Glass |
| Wavelength | 1054 nm |
| Output Energy | 1-2 Joules |
| Pulse Width | 30 nsec |
| Repetition Rate | 1 Hz |
| Beam Divergence | 4 mrad |

Receiver

| | |
|-------------------|-----------------------|
| F - number | 1.33 |
| Aperture Diameter | 15 cm |
| Fields of View | 5, 25, 50 and 75 mrad |
| Detector | Silicon PIN diode |

Scanning System

| | |
|---------------------|---------------|
| Azimuth Range | 3 45° |
| Elevation Range | +90° to -5° |
| Position Resolution | 0.009 degrees |

Data Acquisition

| | |
|--------------------------------|----------------------------|
| Range | 10 km |
| Dynamic Range | 5 orders of magnitude |
| Digital Sample Interval | 31 nsec |
| Digital Sample Resolution | 8 bits |
| No. of Samples per Laser Pulse | 2048 in each of 4 channels |
| Rate | 1 Hz |

Transceiver

| | |
|--------|--------------------------|
| Size | 1.4m H x 0.8m W x 1.2m L |
| Weight | 135 kg |

Auxiliary Equipment

| | |
|--------|---------------------------|
| Size | ~1.4m H x ~2m W x 0.6 m L |
| Weight | 190 kg |

System Power Requirements

| | |
|--|-------|
| | <2 kW |
|--|-------|

SYNCHRONOUS FLUORESCENCE SPECTROSCOPY OF
DISSOLVED ORGANIC MATTER TO
OPTIMIZE LIDAR DETECTION PARAMETERS

Anthony Vodacek, Ph.D. Candidate

Cornell Laboratory for
Environmental Applications of Remote Sensing
Cornell University, Hollister Hall
Ithaca, NY 14853-3501

Abstract

Remote sensing of dissolved organic matter (DOM) in natural waters by laser-induced fluorescence has been limited to emission spectra which are not sufficient for discriminating changes in composition of DOM. Fluorescence quenching by various cations (e.g. hydrogen ion) only adds to the uncertainty. Synchronous fluorescence spectra can potentially provide compositional information since spectral peak positions may indicate DOM composition. The usefulness of synchronous data for predicting pH was tested for 21 lake samples and compared to the results obtained for emission data. Also, predictions of DOC using emission were compared to predictions using synchronous data for 11 of the lake samples. The synchronous data gave a better correlation with pH ($R^2=0.90$) than the emission data ($R^2=0.71$), and matched the high correlation of DOC and emission data ($R^2=0.94$). The critical synchronous data in this example could be reproduced with a three-wavelength lidar.

Key words: remote sensing, fluorescence, synchronous fluorescence, lidar, water, DOM

1. Introduction

Dissolved organic matter (DOM) in natural waters, particularly the humic substance (HS) fraction, is ecologically important as part of the carbon cycle and as a modifier of trace metal chemistry (Wetzel, 1983). Since DOM fluoresces intensely with ultraviolet excitation, DOM has been a frequent target of hydrographic lidar systems, both as the primary target (Hoge and Swift, 1982) and as an interference in the fluorescence spectra of other compounds (Bristow et al., 1981).

DOM fluorescence intensity is usually related to the dissolved organic carbon (DOC) concentration, though attempts have been made to measure hydrogen ion and aluminum concentrations indirectly, by their quenching effect on DOM fluorescence emission spectra (Vertucci and Vodacek, 1985; Philpot and Vodacek, in press). These analytical uses of DOM fluorescence are hampered by unstable DOM composition, since the fluorescence is determined in part by the composition. DOM is a mixture of organic compounds, some nonfluor-

escent, and others fluorescent (e.g. HS), but with varying efficiency and spectral distribution. Unstable DOM composition arises as DOM is altered by both abiotic and biotic processes (Münster, 1985; Ertel et al., 1986). The lack of a fluorescence technique for identifying DOM composition limits the use of remotely sensed DOM fluorescence emission data to cases where stable, conservative behavior of DOM can be shown from grab sample analysis. Therefore, enhancement of DOM fluorescence characteristics that are dependent on the composition of the mixture is a worthwhile goal.

Present theory on fluorescent DOM structure states that it is comprised of single and fused aromatic rings, but with many aliphatic side chains and functional groups substituted on the aromatic rings. It is well known that the extent of coupling of the molecular electronic structure determines the fluorescence wavelength of aromatic hydrocarbons (Inman and Winefordner, 1982). The more coupled the aromatic system, the further to the red the fluorescence emission. Thus, the degree of coupling of the electronic structure can be expected to be responsible for some of the inherent spectral variability between DOM compositional classes. A complicating factor is introduced when ring substitution induces a red shift (Berlman, 1965). Functional groups can shift the fluorescence of a one ring compound to the spectral position normally found for saturated two or three-ring systems. An approach that takes advantage of these phenomena is synchronous fluorescence spectroscopy. This technique employs simultaneous scanning of the excitation and emission with a constant wavelength or energy difference. A synchronous fluorescence spectrum occurs as a diagonal line on a total fluorescence plot of excitation wavelength, emission wavelength, and intensity (Figure 1). The compositional classes will be distributed along the diagonal (the synchronous spectrum), depending on the individual red shifts. Maximum separation of molecular types occurs when small wavelength or energy differences are observed. This technique has been applied to natural DOM samples as a "fingerprint" method for identifying DOM

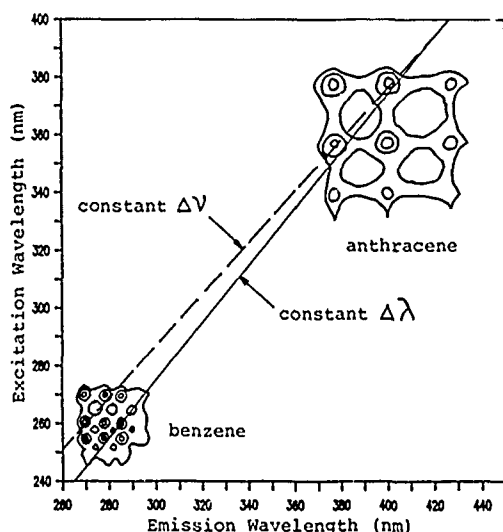


Figure 1. Stylized total fluorescence spectra of benzene and anthracene. Contours represent intensity.

composition (Russo, 1983; Cabaniss and Shuman, 1987).

The effects of trace metals and pH on emission spectra (Vodacek and Philpot, 1987) are also observed for synchronous spectra of natural waters (Cabaniss and Shuman, 1987). Unlike emission spectra, the induced effects can be expected to vary depending on the synchronous peak being observed.

The goal of the study was to use synchronous spectra of natural DOM samples to select a limited number of excitation and emission wavelengths that are particularly sensitive to DOM composition. If fluorescence emission data by selective excitation can categorize DOM composition in natural waters, the viability of remote DOM fluorescence data will be enhanced.

2. Methods

The lake water samples studied were from New York, Michigan, and Wisconsin, and represent a wide range of water chemistry types. The samples had been stored in the dark at 4°C, for up to 21 months. Fresh samples were not collected since the applicability of the analytical technique itself was being tested.

The laser-induced fluorescence emission spectra of the samples were collected and analyzed as described by Philpot and Vodacek (in press). The spectral data were reduced to three parameters, the weighted fluorescence mean wavelength (F_m), the fluorescence intensity at the water Raman band (F_{ram}), and the ratio of the total fluorescence intensity to the fluorescence intensity at the Raman band (F_{rat}).

Synchronous fluorescence spectra of the lake samples were collected with a spectrofluorometer capable of scanning with a constant energy difference between the excitation and emission monochromators. The energy difference of 1600 cm^{-1} was chosen to match the typical vibrational energy spacing, plus the typical Stokes shift, for polycyclic

aromatic hydrocarbons (Inman and Winefordner, 1982). The spectra are plotted as relative intensity versus the emission wavelength. The spectra are corrected for detector and instrument response by monitoring the fluorescence signal of rhodamine dye in the reference channel. This allows direct comparisons of fluorescence at different wavelengths. The emission spectra of two lake water samples in Figure 2a are compared to their synchronous spectra in Figure 2b.

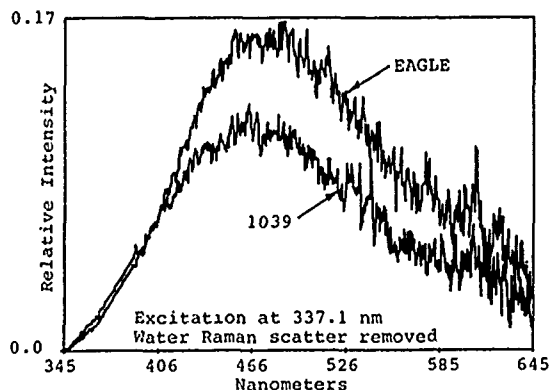


Figure 2a. Emission spectra of two lake water samples.

The synchronous data were reduced to intensities at three wavelengths ($SY1, SY2, SY3$) plus the summed intensity over the whole spectrum (SY_{tot}) as indicated in Figure 2b. Three ratios of the intensities were also calculated ($SY2/SY1, SY3/SY1, SY2/SY3$).

The samples were analyzed for pH after the fluorescence data acquisition. The 11 samples from Michigan and Wisconsin had been analyzed for DOC when the samples were collected.

Multiple linear regression was employed to relate the spectral data (independent variables) to the pH or DOC of the sample (dependent variable). The reported correlation results are for the cases where the coefficients for the independent variables were significant at the 95% level (Student's t-test).

3. Results

Unlike the fluorescence emission spectra, the synchronous spectra of some lakes are very different (Figures 2a and 2b). The peaks in the synchronous spectra can be interpreted as being indicative of different HS composition, possibly modified by pH or trace metal quenching. For DOM emission spectra, the fluorescence of the compositional classes overlaps to a high degree, and correlations with water chemistry are not good. If the intensity of the synchronous peaks is an indication of the prevalence of different compositional units of DOM, the relative intensities can be a simple measure of DOM composition. This partitioning of the compositional classes in the synchronous data might be expected to lead to better correlations with water chemistry than are found for the emission data. To test this, the pH or the DOC

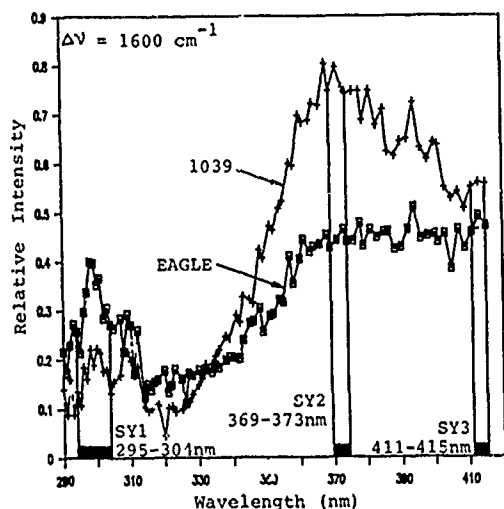


Figure 2b. Constant energy synchronous fluorescence spectra of two lake water samples.

concentration of the samples was predicted using multiple linear regression, first with the emission data, and second with the synchronous data.

The fluorescence emission data for all 21 samples were regressed with the sample pH. The best correlation of the emission data with pH was obtained with the independent variables F_{rat} and F_{Ram} ($R^2 = 0.71$, Std. err. estimate = 0.48 pH units). The emission data for the 11 samples from Michigan and Wisconsin correlated highly with DOC. The significant independent variables in this case were F_m and F_{Ram} ($R^2 = 0.94$; Std. err. estimate = 1.00 mg/L DOC). The plots of predicted versus actual pH and predicted versus actual DOC are given in Figures 3 and 4, respectively.

The synchronous fluorescence data for all 21 samples correlated highly with pH and all independent variables were significant ($R^2 = 0.90$; Std. err. estimate = 0.28 pH units). The synchronous data of the 11 samples from Michigan and Wisconsin gave very similar results to those of the emission data when regressed with DOC. Only four of the synchronous parameters were significant, SY1, SY2, SYtct, and SY2/SY1 ($R^2 = 0.94$; Std. err. estimate = 1.17 mg/L). The plots of predicted versus actual pH and predicted versus actual DOC are given in Figures 5 and 6, respectively.

Although these results must be qualified by the small number of samples, it appears that synchronous spectra are capable of partitioning DOM fluorescence in a way that enhances the effects of pH on the different spectral regions of DOM fluorescence. This implies a link between the synchronous spectra and the DOM composition. However, in this case, synchronous spectra do not seem to contain more useful information than do emission spectra for predicting DOC concentration.

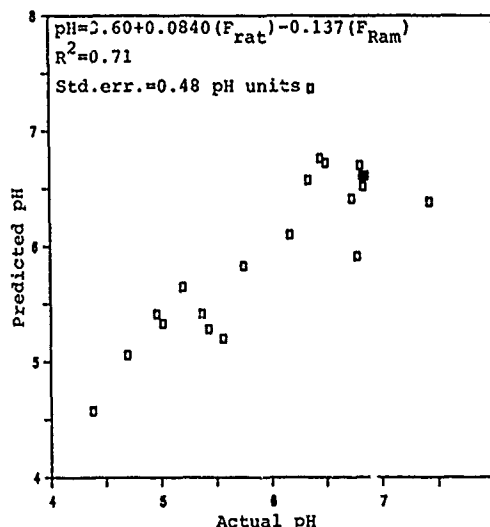


Figure 3. pH predicted by fluorescence emission data versus actual pH.

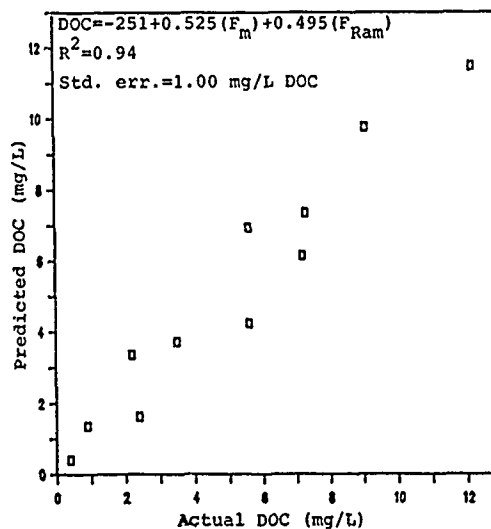


Figure 4. DOC predicted by fluorescence emission data versus actual DOC.

4. Discussion

The lakes studied here exhibited three consistent synchronous fluorescence peaks and these three emission-excitation pairs appear to relate to the DOM composition. While synchronous spectra per se are not readily adaptable to remote sensing because of the need to scan a wide wavelength range, a lidar with a limited number of excitation wavelengths may be sufficient if key excitation-emission pairs can be identified, as in this study. A lidar that could excite at 286 nm, 350 nm, 387 nm, and detect the spectral emission would provide similar data as were obtained from the synchronous spectra. Laser systems capable of multiwavelength excitation have been described by Mumola et al. (1975) and Phinney et al. (1988).

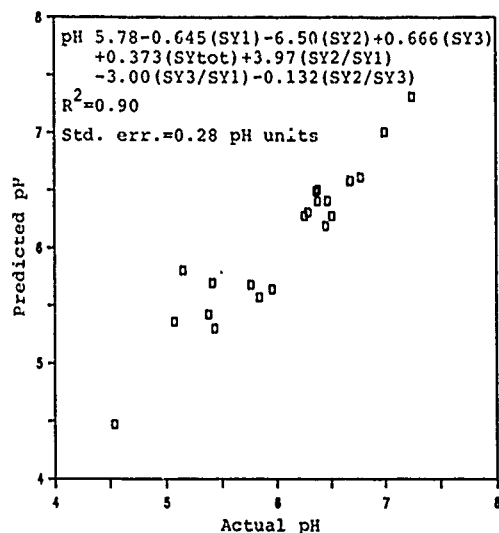


Figure 5. pH predicted by synchronous fluorescence data versus actual pH.

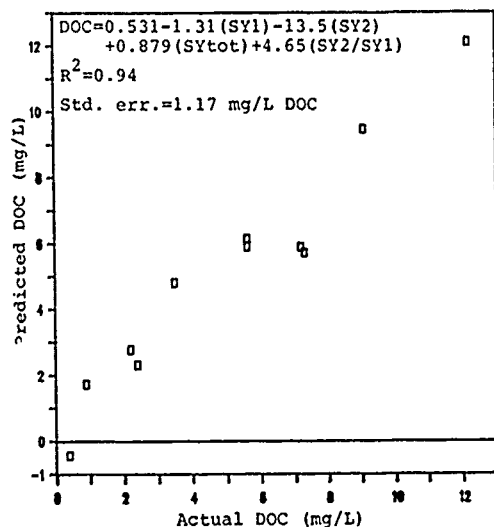


Figure 6. DOC predicted by synchronous fluorescence data versus actual DOC.

Work on multicomponent analysis techniques will continue with the acquisition of DOM total fluorescence spectra and synchronous spectra. Use of total fluorescence may provide a wider range of useful excitation-emission pairs for separation of fluorescent mixtures. If partitioning DOM by fluorescence is viable, the key to its usefulness will be whether the fluorescence data can be transformed into ecologically significant information on DOM. That is, do changes in DOM molecular structure affecting the fluorescence overlap the molecular characteristics affecting the reactivity of DOM in the environment? Future work will employ various physical separation methods to fractionate DOM samples to enable characterization of the fluorescent fractions with total and synchronous fluorescence.

References

1. Wetzel, R.G. "Limnology, 2nd. Ed.", Saunders College Publishing, Philadelphia, 767pp, 1983.
2. Hoge, F.E. and R.N. Wift "Delineation of estuarine fronts in the German Bight using airborne laser-induced water Raman backscatter and fluorescence of water column constituents", *Intl. J. Remote Sensing*, 3:475-495, 1982.
3. Bristow, M., D. Nielsen, D. Bundy and R. Furtek "Use of water Raman emission to correct airborne laser fluorosensor data for effects of water optical attenuation", *Appl. Opt.*, 20:2889-2906, 1981.
4. Vertucci, F.A. and A. Vodacek, "The remote sensing of lake acidification using laser fluorosensing", *Proc.: 51st Annual Meeting, American Society of Photogrammetry*, pp793-801, 1985.
5. Philpot, W.D. and A. Vodacek, "Laser-induced fluorescence: Limits to the remote detection of hydrogen ion, aluminum, and dissolved organic matter", *Remote Sensing Environ.*, in press.
6. Münster, U., "Investigations about structure, distribution and dynamics of different organic substrates in the DOM of lake Plüsee", *Arch. Hydrobiol. /Suppl.*, 70:429-480, 1985.
7. Ertel, J.R., J.I. Hedges, A.H. Devol, J.E. Richey, and M. de Nazaré Goes Ribeiro, "Dissolved humic substances of the Amazon River system", *Limnol. and Oceanogr.*, 31:739-754, 1986.
8. Inman, Jr., E.L. and J.D. Winefordner "Constant energy synchronous fluorescence for analysis of polynuclear aromatic hydrocarbon mixtures", *Anal. Chem.*, 54:2018-2022, 1982.
9. Berlman, I., "Handbook of Fluorescence Spectra of Aromatic Molecules", Academic Press, New York, 258pp, 1965.
10. Russo, M.C., "Evaluation of fluorescence spectroscopy as a method for measuring 'Gelbstoff' in aquatic environments", Ph.D. Thesis, Univ. of London, 1983.
11. Cabaniss, S.E. and M.S. Shuman "Synchronous fluorescence spectra of natural waters: Tracing sources of dissolved organic matter", *Mar. Chem.*, 21:37-50, 1987.
12. Vodacek, A. and W.D. Philpot, "Environmental effects on laser-induced fluorescence spectra of natural waters", *Remote Sensing of Environ.*, 21:83-95, 1987.
13. Mumola, P.B., O. Jarrett, Jr., and C.A. Brown, Jr., "Multiwavelength lidar for remote sensing of chlorophyll a in phytoplankton", *Proc.: The Use of Lasers for Hydrographic Studies*, H.H. Kim and P.T. Ryan, Eds., NASA SP-375, pp137-145, 1975.
14. Phinney, D.A., C.S. Yentsch, and J. Rohrer "Three color laser fluorometer for studies of phytoplankton fluorescence", *SPIE Vol. 925, Ocean Optics IX*, pp171-175, 1988.

Lidar Studies of Atmospheric Extinction

S.R. Pal, A.I. Carswell, A.G. Cunningham

Department of Physics
and
Institute for Space and Terrestrial Science
York University, North York, Ontario, Canada
Tel. No. (416)-736-2100 ext. 7755
FAX No. (416)-736-55.6

ABSTRACT

Measurements of the optical extinction coefficient in the atmosphere have been made using lidar backscattering measurements. This method provides data over an extended altitude range with excellent spatial and temporal resolution. The wide dynamic range of the systems permits measurements of the extinction extending from clear atmosphere into relatively dense clouds. Various approaches have been investigated for deriving the extinction values from the lidar data and these are presented in this paper along with a number of examples of the results obtained under various atmospheric conditions.

KEY WORDS: LIDAR, ATMOSPHERIC EXTINCTION.

The extinction of atmospheric radiation as it propagates in the atmosphere is a phenomenon of considerable importance. The absorption and scattering processes involved govern the deposition and redistribution of solar energy in the atmosphere and place significant limitations on the performance of any optical instrumentation looking through the atmosphere. Lidar, utilizing the range-resolved backscattering of laser radiation is a powerful tool for measuring atmospheric extinction. Not only does lidar provide remote measurements with excellent spatial and temporal resolution but it permits more direct investigation of the fundamental atmospheric absorption and scattering processes than is generally possible with other measurement techniques. Aerosols and clouds are the dominant contributors to extinction in the lower atmosphere and their distributions can be monitored quite readily using fixed frequency lidars. Molecular absorption and inelastic scattering also contribute to the extinction with characteristics which are highly dependent on the particular species and wavelength of incident radiation. Lidars featuring tunable laser sources can be used to selectively measure such species through the differential absorption of two laser pulses whose wavelengths are tuned to different regions of a particular molecular absorption feature. Similarly, lidars featuring a fixed frequency source but wavelength-tunable detection can be used to selectively measure these species through their unique Raman scattering spectral signatures.

Lidar has been used for atmospheric measurements in our laboratory for a number of years during which time several systems were developed for use in a broad range of atmospheric extinction studies. A two-wavelength (694 and 347 nm) ruby lidar and a XeCl excimer (308 nm) lidar have been used to measure extinction in the troposphere using both elastic (Rayleigh, Mie) and Raman scattering.

More recently as part of our newly created Ontario Centre of Excellence, ISTS (Institute for Space and Terrestrial Science), we have built a Nd:YAG based lidar operating at 1064 and 532 nm to extend our measurements into the stratosphere at altitudes up to 60 km. In addition we are developing a Differential Absorption Lidar (DIAL) system for measurements of stratospheric ozone at altitudes between 15 and 35 km.

Prime atmospheric parameters of interest, the extinction coefficient σ (m^{-1}) and the backscatter coefficient β ($\text{m}^{-1} \text{sr}^{-1}$) are related through the lidar equation,

$$P(r) = P_0 \frac{\beta(r)}{r^2} \exp \left[-2 \int_0^r \sigma(r') dr' \right] \quad (1)$$

Where $P(r)$ is the backscatter power from a range r and the constant P_0 includes the transmitted power, receiver area and the instrument calibration factor.

For a quantitative determination of σ this equation can be inverted provided a relationship between σ and β is assumed. Theoretically in a non-absorbing atmosphere they are related through the backscatter phase function, $P_\pi/4\pi$ as, $\beta = (P_\pi/4\pi)\sigma$, however, in a real atmospheric situation they depend in a complex way upon the microstructure of the atmosphere and are spatially and temporally variable. Therefore a relationship, $\beta = C\sigma^k$ (C, k constants) is used to obtain a solution of Eq.1 given by (Klett, 1985)

$$\sigma(r) = \frac{\exp[(S - S_f)/k]}{[\sigma_f^{-1} + \frac{2}{k} \int_f^r \sigma_f \exp[(S - S_f)/k] dr']} \quad (2)$$

where $S(r) = \ln(r^2 P(r))$ and S_f, σ_f, r_f , are the boundary conditions at the far end of the profile.

We have investigated inversion methods involving both modelling and additional measurements of the atmospheric properties to reduce the ambiguity associated with the assumed coupling between σ and β . Various aspects of lidar inversion techniques have been examined using lidar measurements of differing atmospheric conditions in an effort to optimize the accuracy of these inversion solutions. Our analytical procedures yield rather accurate extinction coefficient profiles for the probing of tropospheric clouds where a three order of magnitude variation is often exhibited within a single profile. In clearer atmospheres demonstrating lesser variations in extinction coefficient, the accuracy of the inversion methods is also quite good if care is taken in the estimation of the boundary value for the extinction coefficient. The effects of this boundary value on the inverted extinction profiles has been examined in some detail.

The backscatter profile of Fig. 1 demonstrates the lidar capability to successfully resolve many layers in a multilayer cloud of about 1 km thickness, situated with a 1 km base altitude.

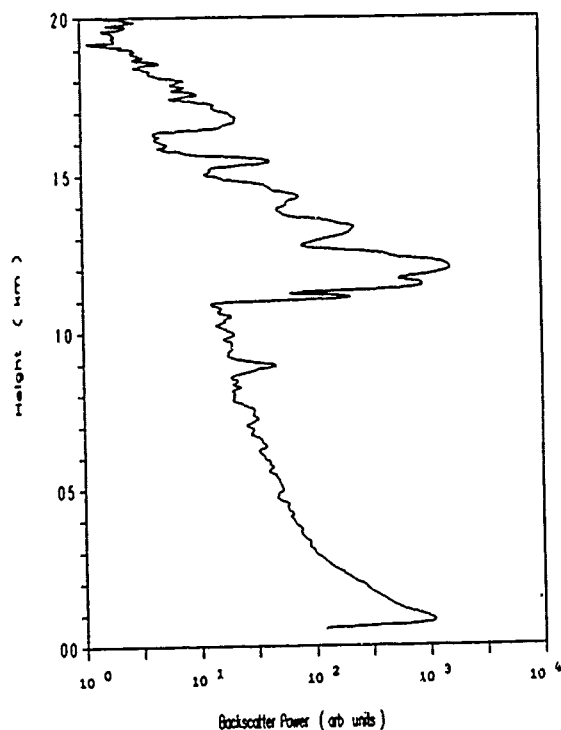


Fig.1: Lidar backscattering profile.

With $\sigma_f = 0.5 \text{ km}^{-1}$ at $r_f = 2 \text{ km}$ a sample inversion of the data in Fig.1 using Eq.2 is given as curve A in Fig.2 where σ is shown as a 5 decade logarithmic plot as a function of height. It has been found that for lidar returns in which clouds are present initial values of greater than 0.5 km^{-1} are appropriate because such values provide little variation in near field value of the extinction coefficient. Using these near field values, σ_0 , further iterations provide improvement on the initial profile as shown by curve B. This approach involves the use of the Newton iteration method to minimize the difference between the measured and calculated values of $S(r)$ on a point by point basis along the profile.

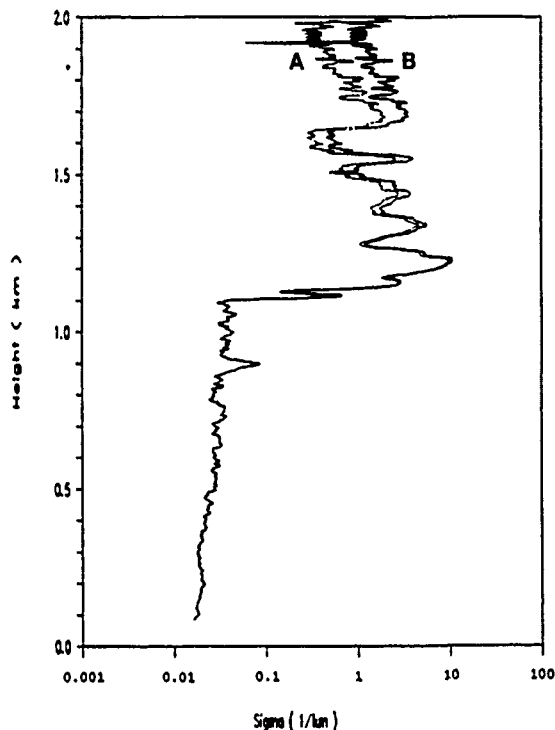


Fig.2 Extinction profiles derived from the data of Fig.1.

Lidar is also capable of demonstrating atmospheric fine structures and their dynamics. The time-height intensity plot of 200 lidar profiles shown in Fig.3 gives a time history of atmospheric attenuation variation over a period of 2 hours.

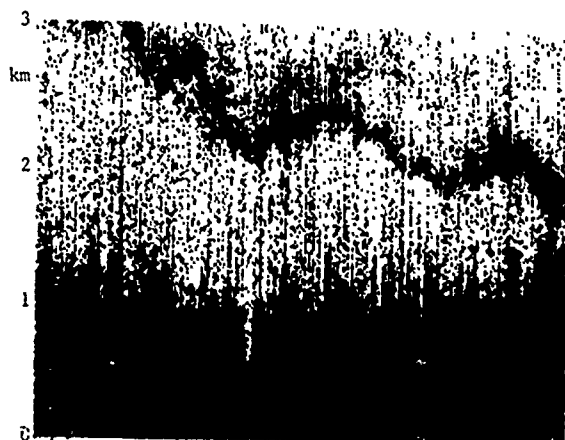


Fig.3: Time-height intensity variation using 200 lidar pulses.

Details of the spatial and temporal variation of the atmospheric extinction are clearly visible in such lidar data. This figure shows an interesting feature at an altitude of about 2 km. This low density aerosol layer is descending during the time of observation with an oscillatory amplitude of about 380 m. Such time series of lidar data can show the three dimensional behaviour of atmospheric extinction with excellent space and time resolution. A number of examples including colour displays will be included in the conference presentation.

Bibliography:

Klett, J.D., "Lidar Inversion with variable backscatter/extinction ratios", Appl. Opt., Vol.24, No.11, pp.1638-16433, 1985.

The NOAA GEOSAT Program: Monitoring Tropical Sea Level With Satellite Altimetry

Robert Cheney
Bruce Douglas
Laury Miller
Russ Agreen
Nancy Doyle

*National Ocean Service, NOAA
Rockville, Maryland 20852
(301) 443-8556*

The U.S. Navy GEOSAT altimeter satellite has operated successfully for more than four years. Tropical Pacific sea level variations, and all wind speed and wave height values from the initial 18-month classified mission have been computed and released. The Exact Repeat Mission (ERM), which began November 8, 1986, has produced nearly three years of additional data which are routinely processed into unclassified Geophysical Data Records (GDR's) by the National Ocean Service of NOAA and distributed to more than 40 institutions worldwide. In addition, GEOSAT sea level data in the tropical Pacific and Indian Oceans are being processed in near-real time (1-month delay) and are distributed in map form as an operational NOAA product.

GEOSAT altimeter data were initially considered by some investigators to be suitable only for mesoscale investigations of sea level because of low accuracy (a few m) orbits and lack of instrumentation on the satellite to correct for water vapor and ionosphere effects. However, radial orbit error affecting studies of sea level variability is confined to very long wavelength (the orbital circumference), water vapor models exist that are adequate for time scales exceeding a few weeks, and a simple ionosphere model has sufficed because of low solar activity so far in the mission. Thus in the tropical Pacific ocean, NOS investigators succeeded in obtaining detailed maps of monthly sea level variation that agree to a few cm with island tide gauges.

Like earlier satellite altimeters, GEOSAT's primary contribution to ocean dynamics is in determination of sea level variability. The GEOSAT data are unique, however, because of their global, long-term coverage. Whereas previous analyses were limited largely to short-term variability statistics, emphasis is now on generation of discrete sea level time series with record lengths of several years. These time series are roughly equivalent to low-pass filtered tide gauge records, where periods of 2 weeks and longer are resolved. From regular grids of these altimetric records, sea level anomaly maps can be produced, and large-scale movements of water can be traced.

We have derived time series of sea level throughout the tropical Pacific Ocean for the first 4 years of the GEOSAT mission (1985-89). A clear relationship is seen between changes in sea level and fluctuations in the large-scale wind field on time scales of weeks to months. This time period is particularly valuable because it includes the 1986-87 El Nino, providing the first complete picture of sea level variability during an El Nino event. The GEOSAT data also document the cold episode, or anti-El Nino conditions that were observed in 1988-89.

Geoid Estimates in the Gulf Stream from GEOSAT Altimetry Data and a Gulfcast Mean Sea Surface

David L. Porter

The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road, Laurel, Maryland 20707-6099 U.S.A.

Scott Glenn and Allan R. Robinson

Harvard University
Cambridge, Massachusetts 02138 U.S.A.

Abstract

Estimates of the absolute dynamic topography of the sea surface were derived from GEOSAT altimetric measurements by employing a 'synthetic' geoid. The 'synthetic' geoid was estimated by subtracting a mean sea surface computed using a Harvard University model of the Gulf Stream from a mean sea surface derived from GEOSAT radar altimeter data. Harvard University's operational Gulf Stream model was used to estimate the 'true' mean oceanographic sea surface. The model assumes quasi-geostrophic baroclinic conditions with a surface-boundary-layer component with higher order physics. Computations of the ocean's absolute dynamic topography derived from GEOSAT compare well with absolute sea surface height measurements estimated from Air Expendable Bathythermographs (AXBTs) which were made along GEOSAT's ground track in the model region (20 km-100 km spacing of AXBT drops) and with sea surface topography as computed from the model. Preliminary comparisons show frontal locations to be within 40 km for placement of the Stream and within 20 km for placement of the rings, and peak topography differences to be less than 10cm RMS.

Keywords. GEOSAT, Model, Gulf Stream, Altimetry

1 Gulf Stream Model

The Harvard University Gulf Stream model [Robinson and Walstad, 1987] is the dynamical component of an operational "Gulfcast" scheme, that provides weekly forecasts of the Gulf Stream and ring positions [Glenn et al., 1987]. This effort has over two years of results. The model uses quasi-geostrophic baroclinic dynamics and can employ a surface-boundary-layer component with higher order physics [Robinson et al., 1988]. The Gulf Stream and its associated rings are strong oceanographic features that are represented by subsurface feature models in nowcasts and initializations. The location of these features is used to form the initial conditions of the model as shown in the contour map of streamlines in Figure 1a. The Gulf Stream axis and ring locations are determined from infrared data, AXBT data and more recently, altimeter data [Robinson et al., 1989]. After the initial conditions of the model are given, a seven day forecast is made from the model with days 3 and 7 of the fore-

cast shown in Figures 1b and 1c. The large meander of the Gulf Stream at 62° W in Figure 1a, which is pinched off to form a warm core ring by day 3 of the forecast, demonstrate the power of the model to represent the phenomena that occur in this energetic current system. From October 7, 1987 until the end of 1988, these operational forecasts of the Gulf Stream were generated weekly.

A mean sea surface, shown in Figure 2, was formed by averaging all of the daily "Gulfcasts" from October 7, 1987 to October 4, 1988. This preliminary research examined other averaging schemes (e.g., averaging just the 52 initializations), but found little difference between them. Future work will investigate the use of matched averages, i.e., only averaging the model results along a specific satellite track when the altimeter is acquiring data. The GEOSAT pass shown in Figure 2 corresponds to the one pass investigated in this paper. For this pass the mean sea surface due to the oceanography was computed by sampling the sea surface height represented in Figure 2 at positions along the pass where altimetric means were computed.

2 GEOSAT Altimeter

GEOSAT, which was designed and built by The Johns Hopkins University Applied Physics Laboratory, became operational for its Exact Repeat Mission on November 8, 1986. The satellite is equipped with a pencil beam nadir looking radar operating at 13.5 GHz. The satellite measures the distance to the sea surface ten times every second from which a one second mean is formed with a total rms noise level of between 5 and 11 cm [LeShack and Sailor, 1988]. The once per second measurements correspond to one observation every 6.7 km along the groundtrack. The satellite takes nominally 100 minutes per orbit and the groundtrack is repeated every 17.05 days, thus there are 22 complete cycles every year. The altimeter data was corrected for ocean tides, earth tides and the effects of the troposphere and ionosphere [Cheney et al., 1987].

To compute the altimetric mean sea surface along the track shown in Figure 2 a reference pass was chosen. The relative orbit error between a new pass and the reference pass due to the inaccuracies in the initial conditions used to compute the satellite's ephemeris is removed by using either a linear or quadratic least squares fit of the difference. This method removes the long wavelength, $O(500 \text{ km})$, relative orbit error and a small oceanographic signal of the same wavelength. Once this procedure is

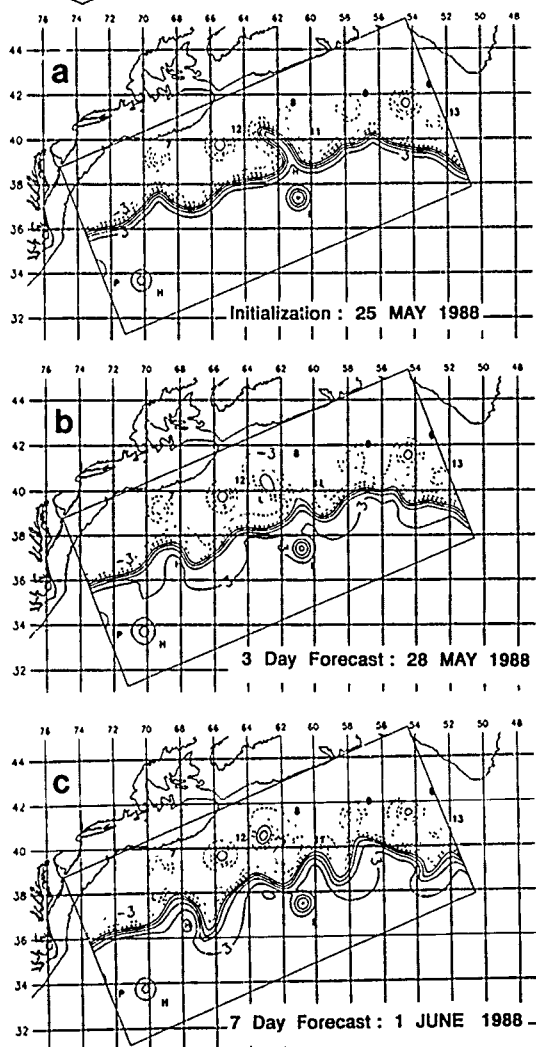


Figure 1 The three panels correspond to (a) the initialization of a Gulf Stream forecast based on IR, AXBT, and GEOSAT data for May 25, 1988, (b) is day 3 of the forecast for May 28, 1988 and (c) is 7 day of the forecast for June 1, 1988. The contour lines are streamlines with the dashed values being less than zero.

carried out for all the passes the mean sea surface can be computed at each observation point along the reference pass. For a year of data that would be an average of 22 measurements at each observation point, however, due to the drop outs of data from the satellite [Cheney et al., 1988], this number is usually less than 22 and for the specific case shown here, it is 20. Thus relative to the reference pass there is a mean sea surface that is composed of the unknown geoid and the average of the oceanographic signal. Additionally there is the unknown orbit error in the reference pass, thus we are not able to absolutely determine a geoid estimate, hence the name 'synthetic' geoid. However, if all future measurements are referenced to the same reference pass they will contain the same unknown orbit error and when the 'synthetic' geoid is subtracted from the new pass to calculate the absolute topography, the unknown orbit error will cancel out.

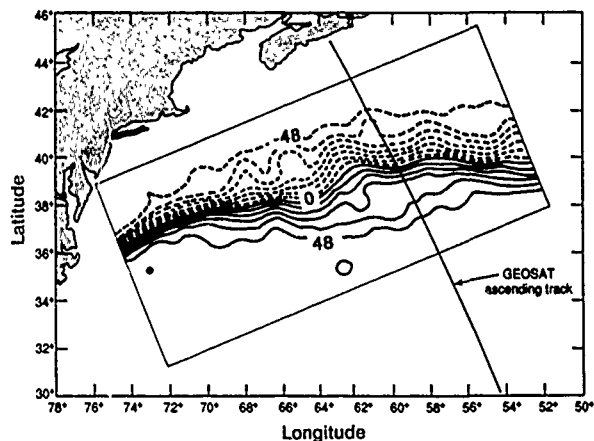


Figure 2 The isopleths are levels of constant sea surface topography in intervals of 8 cms for a one year average of the Gulf Stream model. The line cutting perpendicular to the Stream is the ascending GEOSAT track that was used in this study.

A sensitivity study of the altimetric mean sea surface was made to determine whether a linear or quadratic equation should be used to remove the orbit error, and also whether only altimetric data acquired over the deep ocean i.e. greater than 2250 m, should be used. The orbit error sensitivity was examined first by using the ascending pass shown in Figure 2. The satellite track was 2500 km in length and stopped at the 2250 m bottom contour. A mean sea surface computed using a quadratic fit and a mean sea surface computed using a linear fit to the relative orbit error were differenced and plotted as a function of latitude in Figure 3. Note that if a linear fit is chosen and there is a higher order term in the orbit error, then that term will manifest itself as part of the oceanographic signal. And similarly if the quadratic fit is used it may remove a quadratic term from the oceanographic signal. It is impossible from this preliminary investigation to say which of the two occurs, however, as the difference plot in Figure 3 shows the resultant curve has a small amplitude, 8.4 cm, and a long wavelength, 850 km. Hence, the type of signal that would be wrongly added using the linear or wrongly removed using the quadratic is on the order of 1 cm/100 km. Since the type of signal we are measuring in the Gulf Stream is on the order of 100 cm in 100 km the orbit error removal using either method will not corrupt the measurement.

A similar sensitivity study was performed on the variation in mask depth used. Five depths were used ranging from zero to 2250 m depth with the comparison between these two extremes yielding the largest change in the mean sea surface i.e. 1.5 cm over an arc 860 km in length. This gives a slope of .2 cm/100 km, which is a negligible effect.

3 Comparisons

For this preliminary investigation an arc 2500 km in length terminating at the 2250 m depth contour and a quadratic equation used to remove the orbit error relative to a chosen reference pass were used. For the year of data beginning on October 7, 1987 there were 20 repeat passes over this track that were used to form an altimetric mean sea surface that consists of the geoid, the mean oceanography and some undetermined orbit error associated with the reference pass. From the Harvard Gulfcast estimates of the Gulf Stream a one year average consisting of 364

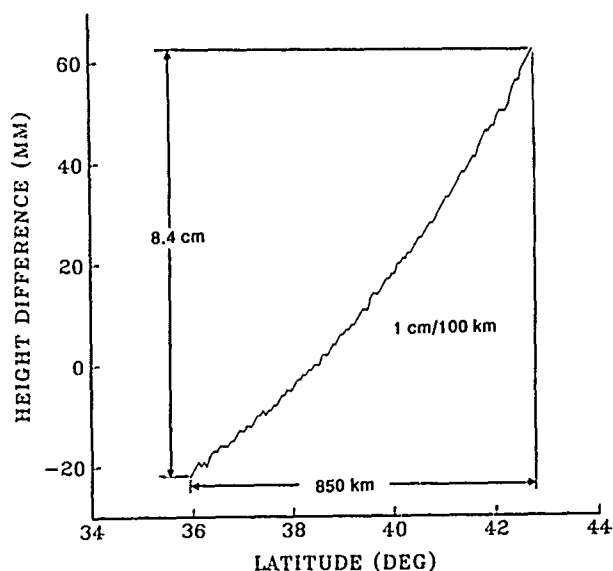


Figure 3 This plot is the difference of the mean sea surface computed using a linear fit to the orbit error minus the mean sea surface computed using a quadratic fit to the orbit error as a function of latitude for the single pass shown in Figure 2.

days of Gulfcast realizations were made and subtracted from the altimetric mean to obtain an estimate of the 'synthetic' geoid.

Figure 4 shows a comparison of the absolute topography computed from the altimeter using the 'synthetic' geoid for October 8, 1987 compared with the model output along the identical ground track for the same day. The amplitude of the Stream is larger for the altimeter than for the model, however the placement of the stream axis and the placement of the warm core ring are within 20 km. The rms error between the two measurements was 22 cm and the correlation coefficient between the two curves .91. This is an encouraging preliminary result and many more passes and comparisons need to be made. We believe that the methodology works well and that the Geosat estimate of dynamic topography is more accurate than that forecast by the Gulfcast scheme. The data compared here was also used in the formation of the sea surface, however similar comparisons between model results and absolute topography can and will be made outside of the period of comparison. As long as the new data is referenced to the same reference pass the 'synthetic' geoid can be used.

An AXBT survey was carried out along the same ground track on May 6, 1987 by the Naval Oceanographic Office. Measurements were made of the oceans temperature as a function of depth by dropping AXBTs at a nominal spacing of 20 km along the GEOSAT groundtrack. The sea surface topography was estimated from the temperature-depth profiles after the method of deWitt [1987]. These observations were made before the time period used to form the means. Since the AXBTs measure only the baroclinic portion of the absolute sea surface height, the values may have a bias added to them. This was done for the two curves shown in Figure 5 where they were forced to agree at the point marked by an 'A'. The solid line is the sea surface height determined from the AXBTs and the dashed curve is the sea surface height determined from the altimeter. Note that the along track placement of the eddy and the Gulf Stream show coincidence. Figure 5 shows a deep cold core eddy located at approximately 37.5° N with a surface depression of about 90 cm and the Gulf Stream located at 40° N with an amplitude of about 90 cm also. The amplitude of the Stream as measured between

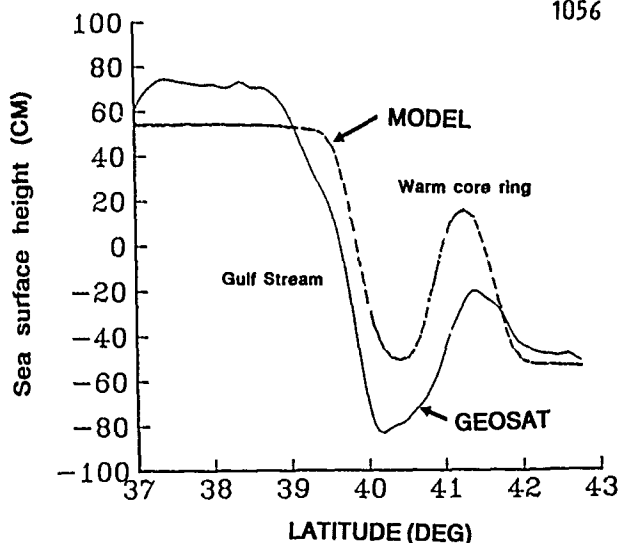


Figure 4 The solid curve represents the absolute sea surface topography computed using the altimetric data and the 'synthetic' geoid for October 8, 1987. The dashed curve is the corresponding realization of the sea surface height based on the 'Gulfcast' for that day. A warm core ring is located at 41.5° N and the main core of the Gulf Stream is located at 39.8° N.

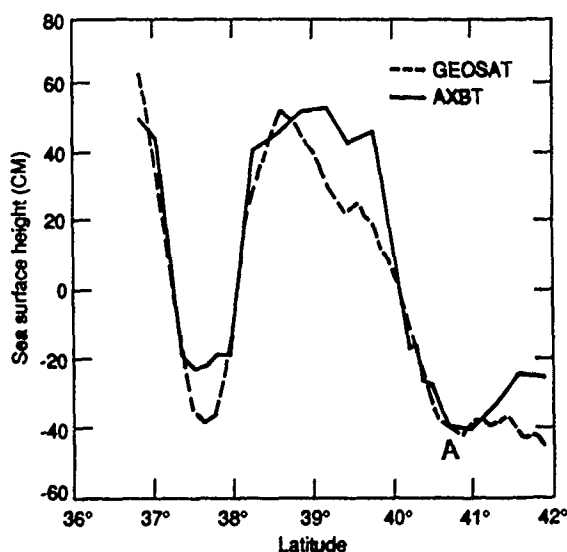


Figure 5 The solid curve is the sea surface height derived from the temperature at 300 m based on the measurements taken from the AXBTs. The dashed line is the absolute topography computed from the GEOSAT altimeter for May 6, 1987. The point marked by an A is where the sea surface heights of the two curves were set equal to each other.

the two extremes of height agree to within 1 cm. The correlation coefficient between the sea surface height derived from the AXBTS and the GEOSAT altimeter is .96 and the rms between the two is 8.8 cm.

4 Conclusions

These preliminary results of our Phase I study show that the computation of the 'synthetic' geoid is a viable method for obtaining the absolute sea surface topography. Once the 'synthetic' geoid is estimated it can be used for all time as long as all new passes are referenced to the same reference pass. The agreement with in situ data collected from AXBTs agrees to within 8.8 cm rms with a correlation coefficient of .96.

Acknowledgements. We gratefully acknowledge the comments from the Harvard JHU/APL working team for their help in formulating this research. We are also grateful to Dr. John Blaha of NAVO who collected the AXBT data. This work was, in large part, supported by the National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) and the JHU/APL GEOSAT program both under U.S. Navy Contract #N00039-87-C-5301 and by the Oceanographer of the Navy by a contract through NORDA to Harvard University.

5 References

- Cheney, R., B. Douglas, R. Agreen, L. Miller, D. Porter, and N. Doyle, "GEOSAT Altimeter Geophysical Data Record User Handbook", NOAA Technical Memorandum NOS NGS-46, U.S. Department of Commerce, Rockville, MD, July, 1987.
- Cheney, R., B. Douglas, R. Agreen, L. Miller, and N. Doyle, "The NOAA GEOSAT Geophysical Data Records: Summary of the First Year of the Exact Repeat Mission", NOAA Technical Memorandum NOS NGS-48, U.S. Department of Commerce, Rockville, MD, September, 1988.
- deWitt, P. W. "Subsurface Temperature Structure as inferred from Sea- Surface Topography - A Possible Application of Satellite Altimetry", Technical Report TR-295, Naval Oceanographic Office, October 1986.
- Glenn, S., A. Robinson and M. Spall, "Recent Results from the Harvard Gulf Stream Forecasting Program", Oceanographic Monthly Summary, 7, No. 4, April, 1987.
- LeShack, A.R., and R.V. Sailor, "A Preliminary Model for Geosat Altimeter Data Error", GRL, Vol. 15, No. 11, pp 1203-1206, October, 1988.
- Robinson, A.R., M.A. Spall, and N. Pinardi, 1988, "Gulf Stream Simulations and the Dynamics of Ring and Meander Process", JPO, 18, No. 12, pp. 1811-1853.
- Robinson, A.R. and L.J. Walstad, 1987, "The Harvard Open Ocean Model: Calibration and Application to Dynamical Process, Forecasting, and Data Assimilation Studies", Applied Numerical Mathematics, 3, 69-131.
- Robinson, A.R., M.A. Spall, L.J. Walstad, and W.G. Leslie, "Data Assimilation and Dynamical Interpolation in GULFCAST-ING Experiment", to appear in Dynamics of Atmospheres and Oceans, 1989.

SEA LEVEL VARIATIONS OF THE N.E. ATLANTIC MEASURED BY GEOSAT

JEREMY THOMAS
 SCHOOL OF OCEAN SCIENCES UNIVERSITY COLLEGE OF NORTH WALES
 MENAI BRIDGE, GWYNEDD, LL59 5ES, U.K.
 TELEPHONE (0248) 351151
 FAX (0248) 716367
 TELEX 61100
 TELEMAIL UCNW.OCEAN

PHILIP WOODWORTH
 PROUDMAN OCEANOGRAPHIC LABORATORY
 BIDSTON OBSERVATORY, BIRKENHEAD, MERSEYSIDE, L43 7RA, U.K.
 TELEPHONE (051) 653 8633
 FAX (051) 653 6269
 TELEX 628591 OCEAN B

Measurements of sea surface height made by the radar altimeter aboard the U.S. Navy's GEOSAT now cover more than two years, from the beginning of the unclassified data period in November 1986. The 17-day repeat period of the GEOSAT orbit results in a spatial coverage ideal for sea level variability studies of the North East Atlantic and its neighbouring shelf seas.

A number of aspects of the data have been studied, including the sea level variability over the whole region and more detailed variations along sub satellite tracks near to the British coast. The Geophysical Data Records (GDR) for GEOSAT, supplied by NOAA/John's Hopkins University Applied Physics Lab., include correction factors for all the geophysical variables which introduce delays or errors into the radar pulse travel time. Predicted sea level variations due to solid earth and ocean tides are also supplied. All these correction factors have been used in this study, except for the ocean tide predictions from the Schwiderski global ocean tide model. These have been replaced by a finer resolution model of the N.E. Atlantic developed by Flather at the Proudman Oceanographic Lab.. This has significant impact on the results for North West European shelf seas.

Maps of root mean square sea level variability for the whole region have been produced, using repeat track analysis to remove orbit determination errors. Significant spatially and temporally coherent oceanic features can be seen, despite the smaller amplitude of variations compared to a western boundary current region, for example. Summer and winter variability maps have been produced to investigate seasonal differences. Also the impact of the significant wave height correction and the inverse barometer effect of atmospheric pressure, on such maps have been studied.

For individual track investigations, advantage has been taken of the large database of tide gauge measurements available at the Proudman Oceanographic Lab., especially where they lie close to sub-satellite points. For example the tide gauges at Lerwick in the Shetland Isles and at Wick in northern Scotland lie near the same track. This enables comparison of height differences along the satellite track with height differences between the tide gauges. These height differences are shown to be highly correlated before the ocean tide signal is removed from each. This indicates that the GEOSAT altimeter is capable of measuring ocean tide

differences to good accuracy and therefore the data can be used to derive significantly improved ocean tide models. A technique for removing the ocean tide signal on scales too small for the Flather model to resolve has been tested. The resulting residual height differences show a fair degree of agreement between altimeter and tide gauges, especially where certain events causing decimetric sea level differences have occurred. These are being investigated using a storm surge prediction model and in-situ current meter data.

The aspects of this study investigated and the results obtained so far have been summarised here. Further work is being undertaken, especially with regard to storm surge model comparisons and combination of tide gauge and current meter data with altimetry, to monitor events causing sea level differences in North West European shelf seas.

Sea-surface height anomalies in the north-east Pacific as observed with the Geosat altimeter: the Sitka eddy

J F R Gower

Institute of Ocean Sciences,
P.O. Box 6000, Sidney, BC, V8L 4B2, Canada
Tel: (604) 356-6558, Fax: (604) 356-6479, Telex: 049-7281

ABSTRACT

Data from the first two years of the Exact Repeat Mission of Geosat have been analysed to detect sea surface height anomalies in the north-east Pacific (45° to 60°N and 120° to 150°W). Consistent anomalies with space scales of order 100–300 km, recurring after successive 17 day intervals, are observed on individual satellite passes with amplitudes of 10 to 30 cm, and with anomaly displacement velocities along the satellite track of up to about 1 cm/s. One of the major features in the Geosat data was previously identified from ship measurements and named the "Sitka" eddy. The Geosat data give improved temporal and spatial information on such eddies, and are used here to study the persistence and intensity variations of this feature. Satellite altimeter data should be able to provide near-real-time information on such eddies and on their associated surface currents.

Key words: Satellite altimetry, Geosat, ocean circulation, eddies.

GEOSAT ALTIMETRY DATA

The US Navy Geosat satellite has provided altimetry data in its Exact Repeat Mission (ERM) since November 1986. The altimeter measures the travel time of pulses emitted vertically downwards, and reflected back to the satellite by an area of the earth's surface typically a few kilometers across. Accurate knowledge of the satellite's position and of the pulses' propagation speed, allows measurement of the absolute height of the earth's surface. Over the oceans, the mean flatness of the sea surface on kilometer scales, and the symmetry of height offsets introduced by surface waves, allows a mean range to be determined with a precision of a few centimeters.

Such a precision gives important oceanographic data on ocean circulation and on smaller scale motions, since the geostrophic balance of these water movements on a rotating earth requires surface height changes in the range 10 to 100 cm. Oceanic applications of such data have been demonstrated by Cheney et al. (1983), Fu and Chelton (1985) and Tai et al. (1989).

The Exact Repeat Mission was chosen to duplicate the coverage pattern provided by Seasat in 1978. In the repeat period of 17 days, 244 satellite orbits cover an oblique grid with an equator crossing separation of 164 km. The coverage pattern is defined by ascending node longitudes $1.05^{\circ}\text{E} + 360n/244$ for $n = 1$ to 244, orbit inclination of 108.04° and period of 6037.55s.

OBSERVATIONS IN THE N.E. PACIFIC

The present study covers the area 45 to 60°N , 120 to 150°W . Ascending orbit tracks cross this area with a mean spacing of about 90 km as shown in Figure 1. These tracks correspond to $n = 166$ to 173 in the formula for ascending node longitudes given above. Similar descending tracks also cross this area approximately orthogonally to those shown. Much of the data from these tracks are unusable due to an attitude instability of the satellite after crossing arctic Canada, and information from these orbits is not considered here.

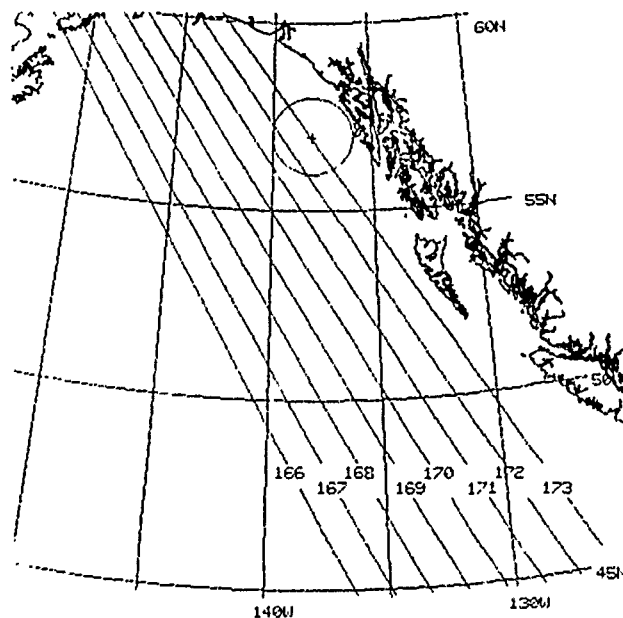


Figure 1. Map of the study area showing the coast of Alaska and British Columbia, the positions of Geosat ascending orbit tracks for $n = 166$ to 173 , and the position and approximate scale of the recurring "Sitka" eddy reported by Tabata (1982).

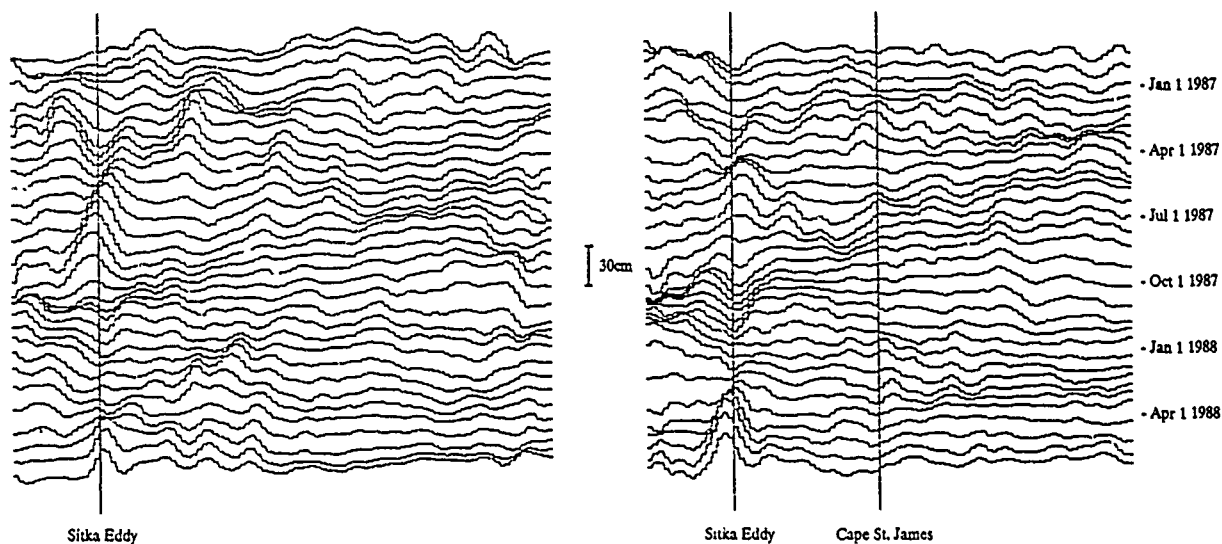


Figure 2a. Plots of anomaly heights deduced from Geosat data, relative to the mean of the first 34 repeats in the ERM, for tracks corresponding to $n = 172$ (left) and 173 (right). Data are shown for the tracks plotted in Figure 1, with latitude decreasing to the right. The vertical line indicates the latitude of the Sitka eddy reported by Tabata (1982). The latitude of Cape St. James is also shown for reference. A vertical scale of 30cm is shown at the centre. Approximate dates are indicated to the right.

Copies of tapes of the Geosat global data set are received from NOAA in the USA, via MEDS in Ottawa, typically 6 months after data collection. One 6250 bpi tape covers two 17 day cycles. Corrections listed on the tapes for variations in geoid height, solid earth and ocean tides, water vapour and atmospheric pressure (propagation and inverse barometer corrections), ionosphere electron content and satellite attitude, are applied to the altimetry data.

The major error remaining is in knowledge of the satellite's absolute vertical position. This can amount to several meters, but is slowly varying over a spatial scale of several thousand kilometers. The error for each orbit track is compensated in the present study by subtracting the best fit second order polynomial function of along-track distance from the observed surface heights. The correction is computed separately for each satellite pass along each track over deep water in a larger area covering 0° to 60°N and 90° to 180°W .

Inaccuracy of the geoid heights listed on the Geosat data tapes is also significant at the space scales ($<500\text{km}$) being studied here. Height anomalies are therefore computed as the difference of the corrected heights along each track from the two year mean for that track.

Repeat measurements of sea surface height anomalies computed in this way at 17 day intervals from November 1986 to June 1988 are displayed in Figure 2. Data from each satellite track are plotted along a horizontal axis, with the 34 repeat measurements made by Geosat in the above period, aligned vertically by latitude. Missed passes (15 out of 272 discussed here) and about 2% of the data which were manually deleted have been linearly interpolated. The data were then smoothed over 60 consecutive kilometers along track and over 3 consecutive passes between tracks. In Figure 2a the data for the 34 repeat passes for $n = 172$ and 173 are plotted graphically. In Figure 2b the height anomalies are coded as grey scale variations from black (about 30cm low) to white (about 30cm high), relative to a mean level shown as a neutral grey. Data are shown for all the passes whose tracks are plotted in Figure 1.

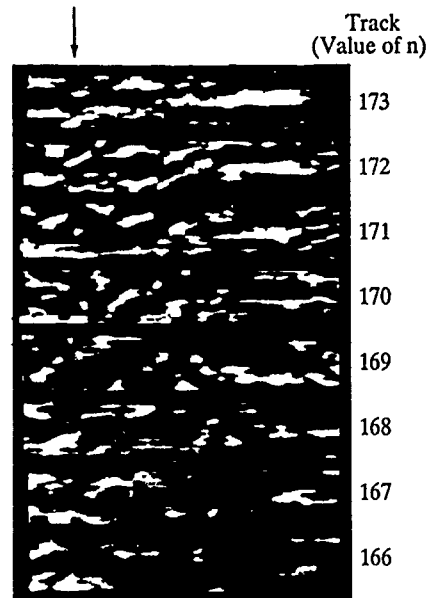


Figure 2b. Grey scale image showing the same data, on the same axes, as Figure 2a, but including all tracks from $n = 173$ (top) to 166 (bottom). Each rectangle shows the height anomaly data for 34 consecutive satellite cycles. Heights are coded as image brightnesses for a range of $\pm 30\text{cm}$ about a mean value. The position of the Sitka eddy, visible on tracks 172 and 173, is indicated by the vertical arrow above the Figure.

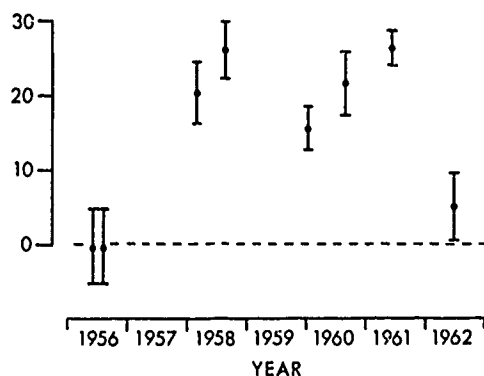


Figure 3. Anomaly heights (in cm) observed for the Sitka eddy in each of the research cruises between 1956 and 1962 reported by Tabata (1982).

The data in Figure 2 cover a time period of 1.5 years. Height variations with an annual period are visible on the left end of each track in Figure 2b corresponding to positions close to the Alaska coast, and also towards the right hand end of the top 3 passes, off the coast of Vancouver Island. These larger scale phenomena may be artifacts of the tidal model used to correct the Geosat data, but are being investigated. They are not discussed further here.

The data clearly show height anomalies with coherence scales of order 150km (along the tracks shown) in space, and 3 months (about 5 repeat cycles) in time. The small scale anomalies are more intense closer to the Alaska/BC coast ($n = 171-173$) than in passes further offshore, and are also concentrated to the left (north-west) ends of the tracks.

Several of the anomalies appear in the same place for periods of a few months. In other cases, movement of the anomaly along the satellite track is evident in both Figures 2a and 2b. A feature showing a trend at 45 degrees to the vertical in Figure 2b would have a speed of 0.6cm/s. Along-track speeds appear to be in the

range 0 to 1 cm/s, with all apparent movement to the north-west. Data presentation is presently being improved to show the consistency of feature visibility from track to track and to study the full motion in two dimensions.

Previous studies of the height anomalies in this area have been reported by Tabata (1982) who shows a series of maps of geopotential anomaly relative to 1000m deduced from hydrographic (bottle cast) station data measured on US/Canada ship cruises between 1956 and 1962. These show height anomalies with similar space and time characteristics to those shown in Figure 2. Stations covered an irregular grid with a spacing of 100 - 200km, concentrating on the area under the satellite tracks in Figure 1, but extending over the whole area shown. Typical cruise duration was about 30 days. The Geosat data provide improved spatial coverage (lines separated by 90km), with observations along the lines spaced at 3 day intervals over the 17 day satellite repeat period.

INTENSITY OF THE SITKA EDDY

The most active eddy reported by Tabata (1982) appeared in about 50% of observations at a roughly constant position off Sitka, Alaska, centered at about 57°N , 138°W , and with a diameter of 200-300km. The approximate surface height anomalies observed on each cruise for this eddy are re-plotted from Tabata's original data in Figure 3 (Tabata, pers. comm.). These show a positive height anomaly of 15 to 30cm, present at this location on over 50% of the cruises.

The Geosat altimeter observations cross the region of this eddy as shown in Figure 1, and the above mean position of the eddy is indicated on Figure 2b. A positive height anomaly of about 15cm occurs at this position in May/June 1987, with a spatial extent of 100km. A larger anomaly of about 20-30cm occurs in April/May 1988.

A constant height anomaly in a fixed position will be removed from the data shown in Figure 2 when the mean observed height is subtracted from the observations. Heights at the expected position of the Sitka eddy in Figure 2 are both positive and negative with respect to the mean, suggesting either a reversing eddy, or variable intensity in one sense of rotation. Tabata's

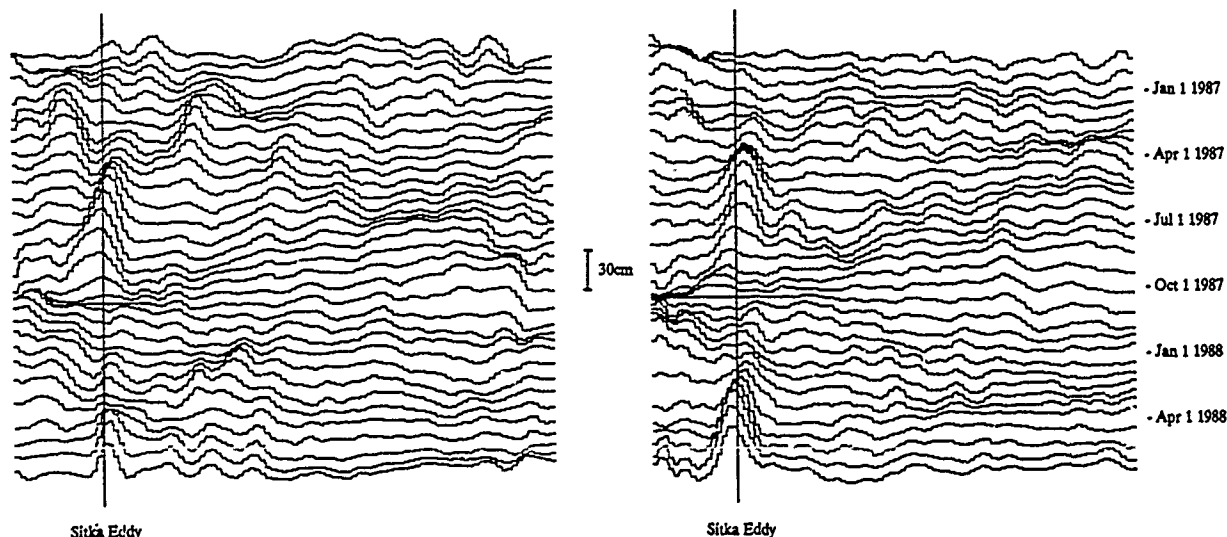


Figure 4. Anomaly heights above the minimum observed for the Sitka eddy. Data presentation similar to Figure 2a.

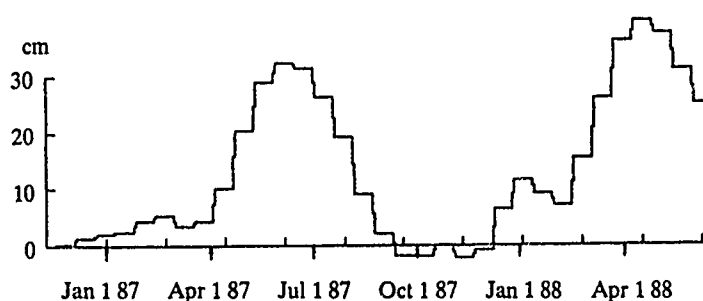


Figure 5. Time series of Sitka eddy anomaly height (in cm) computed from the Geosat data for the track with $n = 173$ for the period November 1986 to June 1988.

observations are of a positive height anomaly, suggesting that negative observed heights are an artifact of the mean subtraction. If the negative anomaly observed in September/October 1987 is interpreted as representing absence of the eddy, then the heights observed at this time should be added to all other observations to produce the revised plot (Figure 4). This naturally shows an intensified eddy at the same location.

The time series of the surface height anomaly is shown in Figure 5. The heights were computed above a linear base line connecting regions 200km on either side of the marked position in Figure 4. The subtraction of the negative (Sept/Oct 1987) anomaly has had the effect of shifting the baseline of this plot downwards.

With this interpretation, the eddy appears to have had two periods of activity during the period shown. One from May to August 1987, the other commencing in February 1988. This type of variability would be aliased by the cruise sampling in Figure 3.

The slopes indicated in Figure 2 in the neighbourhood of the Sitka eddy are of the order of 30 cm over 100km, implying a geostrophic current of about 25cm/s. Such an eddy at the location shown in Figure 1 would induce a south-east flowing surface counter current between latitudes 56 and 58N offshore of the north-westward coastal current, and would therefore have an important effect on the surface flow in the Gulf of Alaska. Such a reversal of the normal flow pattern was of particular concern during the recent Valdez oil spill.

Satellite altimeter data, if available with sufficiently small delay, could therefore be useful for guiding those involved in prediction of oil-spill movements. The present data reception and pre-processing system operated by the applied Physics Laboratory (Johns Hopkins University) and NOAA, can make the data available in about 10 days (L. Miller, pers. comm.), sufficiently fast to influence response planning.

CONCLUSIONS

Even on a relatively local scale, satellite altimetry provides data that can considerably expand knowledge of ocean dynamics. Comparison with ship data previously available, shows the advantage of the satellite's wide and repeated spatial and temporal coverage. Interpretation of the satellite data could have been improved if the descending satellite passes had provided more consistent coverage.

The example of the Sitka eddy, shown above, demonstrates an application of satellite altimetry where near-real-time availability of the data can be important.

ACKNOWLEDGEMENT

Funding from the Radar Data Development Program, administered by the Department of Energy, Mines and Resources, is gratefully acknowledged.

REFERENCES

- Cheney, R.E., J.G. Marsh and B.D. Beckley, 1983, "Global mesoscale variability from collinear tracks of SEASAT altimeter data" *Journal of Geophys. Research*, **88**, 4343-4354.
- Fu, L.-L., and D.B. Chelton, 1985, "Observing large-scale temporal variability of ocean currents by satellite altimetry: with reference to the Antarctic Circumpolar Current", *Journal of Geophys. Research*, **90**, 4721-4739.
- Tabata, S. 1982 "The anticyclonic, baroclinic eddy off Sitka, Alaska, in the northeast Pacific Ocean" *Journal of Physical Oceanography*, **12**, 1260-1282.
- Tai, C.-K., W.B. White and S.E. Pazan, 1989, "Geosat cross-over analysis in the tropical Pacific. 2. Verification analysis of altimetric sea level maps with expendable bathythermograph and island sea level data", *Journal of Geophys. Research*, **94**, 897-908.

Mesoscale variability off California as seen by the GEOSAT altimeter

P. FLAMENT (Department of Oceanography, University of Hawaii at Manoa, Honolulu HI 96822)

P.M. KOSRO and A. HUYER (College of Oceanography, Oregon State University, Corvallis OR 97331)

Abstract. GEOSAT altimeter data was used to study the seasonality, spatial distribution and propagation of mesoscale features in the California Current. Relative topography from the altimeter was first compared with dynamic topography from collinear hydrographic sections. After removing a mean offset, the rms difference was 3-4 cm, barely more than the noise of the altimeter. Statistics of relative topography variability were then estimated over 19 months. Variability reaches 10-12 cm rms near the coast off Northern California, Point Conception, and Point Eugenia. High variability off Northern California extends offshore as two bands ~200 km wide, one towards the southwest, the other one towards the west following the Mendocino escarpment. It decreases westward reaching the background of 4 cm rms at 600 km from the coast. It also decreases to ~4 cm rms north of 42°N and south of 26°S, even near the coast. Time-latitude correlations show poleward propagation of sea level anomalies at about 4 km/day for the two tracks closest to shore, but not further offshore. Rms alongshore topography between 30°N and 40°N has a marked seasonal cycle. Close to shore, it is smallest in fall and winter, and largest during the upwelling season. The cycle lags by ~0.5 day/km and decreases in amplitude westward; it is not detected further than ~500 km offshore. This suggests that mesoscale variability in the California Current is dominated by the growth and seaward migration of meanders of the upwelling front, dynamically linked to the wind-driven coastal upwelling cycle. (California Current, mesoscale eddies, altimeter).

1. Introduction

The summertime mesoscale flow off Central and Northern California consists of large meanders of the California Current. The seaward branches of the meanders are ~40 km-wide baroclinic jets that transport about $2 \cdot 10^6 \text{ m}^3/\text{s}$ of cold water westward from the coastal upwelling region. They appear on satellite infrared images as cold filaments that extend up to 400 km offshore and are generally referred to as upwelling filaments (Brink, 1983, 1987). The structure of these meanders in spring and summer is well known off Point Arena (39°N) and off Point Conception (34°N), having been studied in numerous field programs (Rienecker *et al.*, 1985; Flament *et al.*, 1985; Kosro and Huyer, 1986; Barth and Brink, 1987; Coastal Transition Zone Group, 1988). However, little is known of the structure of the flow at other locations, and at other times of the year.

Our objective in this paper is to study the spatial distribution, the propagation and the seasonality of these mesoscale features along the West coast of North America, from 20°N to 50°N, using ocean surface topography from the GEOSAT altimeter. Because neither the orbit of GEOSAT nor the absolute geoid currently are known with sufficient accuracy, only a variability study can be made in low energy regions such as this one: the mean flow is not presently accessible.

The data and the processing steps followed are described in section 2. The altimeter-derived ocean topography is compared with two collinear hydrographic sections in section 3. The spatial distribution of relative topography variability is presented in section 4 and the temporal distribution in section 5. Possible sources of errors are addressed in section 6.

2. Data and processing

The GEOSAT Exact Repeat Mission provides repeated collinear profiles of sea surface elevation every 17 days with along-track resolution of 7.3 km. The ascending tracks are nearly parallel to the coast of California, and, at the latitude of Point Arena, are found at 85, 215, 345, 475, ... km offshore (Fig. 1). Only ascending tracks were processed for this study because many cycles of the descending tracks are missing (Nov. 1986 to Feb. 1987 and Nov. 1987 to Feb. 1988), making it difficult to resolve the seasonal upwelling cycle unambiguously.

The data were processed for the period November 1986 to April 1988 (19 months, 34 repeat cycles). One-second averaged altimeter heights were corrected for ocean and earth tides, water vapor (FNOC), tropospheric (FNOC) and ionospheric delays, and surface pressure (FNOC) using the corrections given on the NOAA tapes. Orbit errors were modeled as parabolas over 30° arcs and removed. Each profile was then linearly interpolated to a common grid and the average geoid was subtracted to obtain relative surface topography $h(\lambda, t)$, where λ is the latitude and t is the time of each repeat cycle. This procedure removes both the mean and any ~30°-scale oceanic signals.

The along-track power spectrum of relative topography for the track closest to the coast is shown in Fig. 2, normalized by a 3-cm rms gaussian white noise. It is red at wavenumbers smaller than $1/50 \text{ km}^{-1}$ but has a white shape due to instrument noise at higher wavenumbers (the drop of power near the high wavenumber end is caused by the interpolation, cf. Fu, 1983). Because noise dominates at short wavelengths, relative topography was filtered by a running mean over 7 samples (50 km). The contribution of wavelengths longer than 50 km to noise is ~1.2 cm rms.

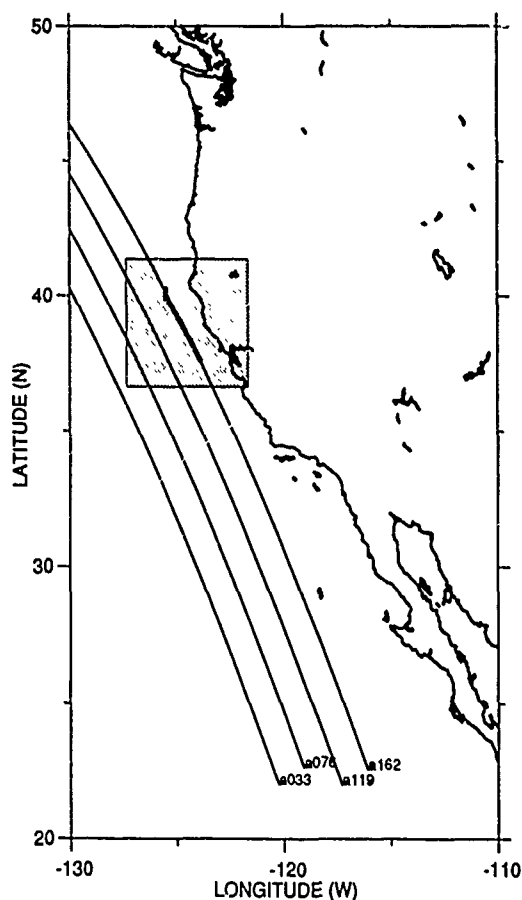


Fig. 1. Map of the area studied, showing the four GEOSAT ascending tracks nearest shore: a162, a119, a076 and a033. The position of the satellite image shown in Fig. 4 is outlined.

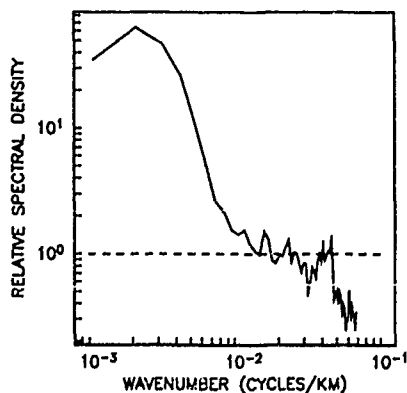


Fig. 2. Along-track power spectrum of relative topography for track a162, normalized by a 3-cm white noise.

3 Comparison with hydrography

A line of CTD hydrographic stations surveyed in May and June 1987 (Schramm *et al.* 1988a, 1988b) coincided with the ascending track closest to the coast. The stations were spaced ~15 km apart. The dynamic topography referenced to 500 dbar was computed and, for each survey, the mean over the length of the section was subtracted. The unfiltered altimeter topography was then interpolated in space and time to the positions of the stations, and, for each survey, the mean over the length of the section was also subtracted.

These relative surface topographies are shown in Fig. 3. In May and June, the altimeter clearly shows the 20 cm signal due to the meander of the California Current seen near 39°N in the AVHRR image of 16 June at 2300 UT (Fig. 4). The topography corresponds to an offshore speed of about ~0.50 m/s over a width of ~30 km. The rms differences between the relative topographies determined from the altimeter and from the CTD sections are 3.3 and 4.2 cm for the May and June surveys. The two topographies are in remarkable agreement, given the ~3-cm rms noise of the radar altimeter (Cheney *et al.*, 1989; Sailor and LeSchack, 1987).

These results show that the altimeter topography is capable of resolving the mesoscale structure of the California Current. They also indicate that, over the period analyzed, the meander was intermittent (if it were a steady feature, its surface expression would be removed with the mean and would not appear in the relative topography), and that the barotropic component was small at this spatial scale, the hydrographic reference level of 500dbar capturing most of the flow (Kosro and Huyer, 1986).

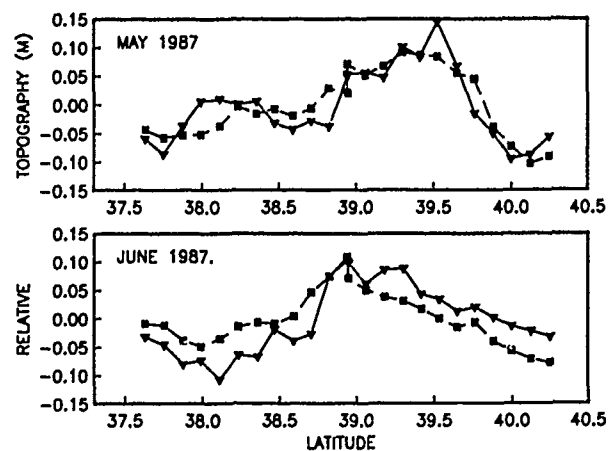


Fig. 3. Comparison between relative topography from the altimeter (dotted line, \square) and from hydrography (solid line, ∇); see text.

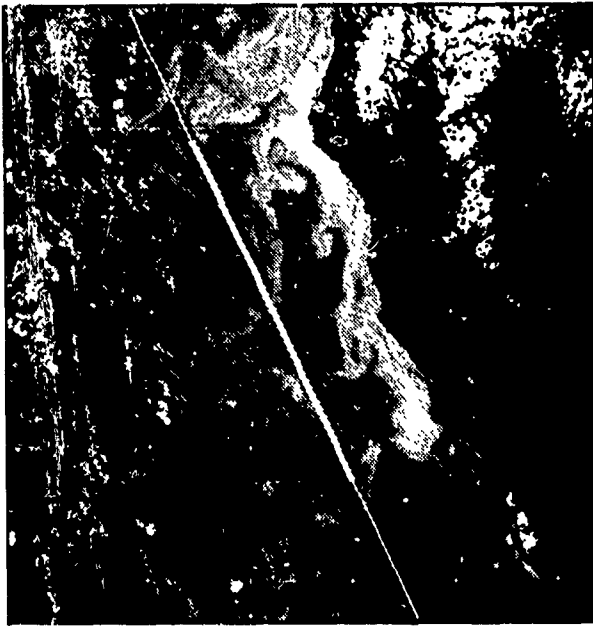


Fig. 4. Thermal infrared AVHRR image on 16 June 1987 at 23:00 UT. Cold water is coded in white and warm water in black. GEOSAT track a162 is shown, with the segment coinciding with the CTD surveys as a thick line.

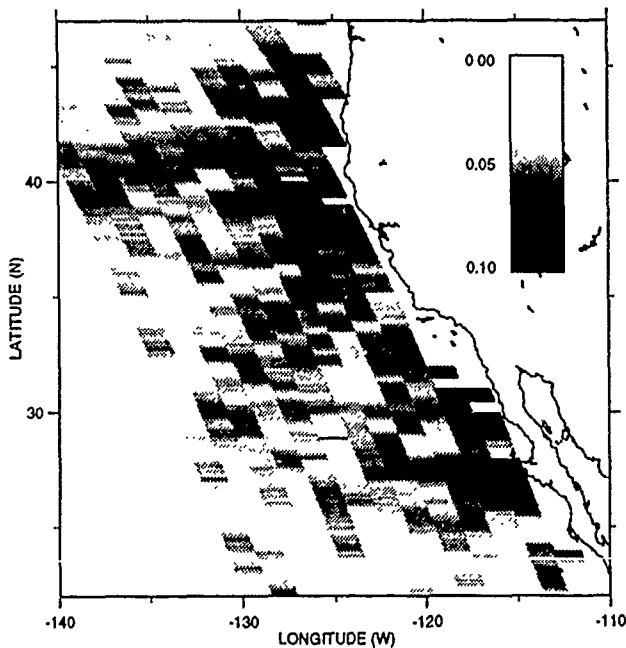


Fig. 5. Map of variability of relative topography over the Eastern Pacific, gray-coded from white (0.00 m rms) to black (0.10 m rms).

4. Spatial distribution of variability

Variability of relative topography, defined as $T^{-1} \int_0^T dt h(\lambda, t)^2$, was computed over the Eastern Pacific for the first 18 months of the mission, and is displayed in Fig. 5 by assigning gray-coded values of variability to rectangular pixels 7.3 km in latitude by 130 km in longitude. Variability along the three ascending tracks closest to the coast is also plotted as function of latitude in Fig. 6. Based on the sample size of 34, the 90% confidence limits are (7,10) cm for a variability of 8 cm rms.

Marked enhancements of variability are seen near the coast of California, reaching 10–12 cm rms west of Point Reyes/Point Arena/Cape Mendocino, south of Point Conception, and around Point Eugenia (Baja California). Variability decreases progressively towards offshore to a typical background value of 4 cm rms at ~600 km from the coast. Variability decreases rapidly to this background value north of 42°N and south of 26°S, even near the coast. Spatial patterns of variability are apparent. The region of high variability off Northern California extends offshore as two bands ~200 km wide, one towards the west following nearly exactly Mendocino escarpment at 40°N, the other one towards the southwest. This region is separated from the one south of Point Conception by a band of low variability, starting surprisingly close to the coast.

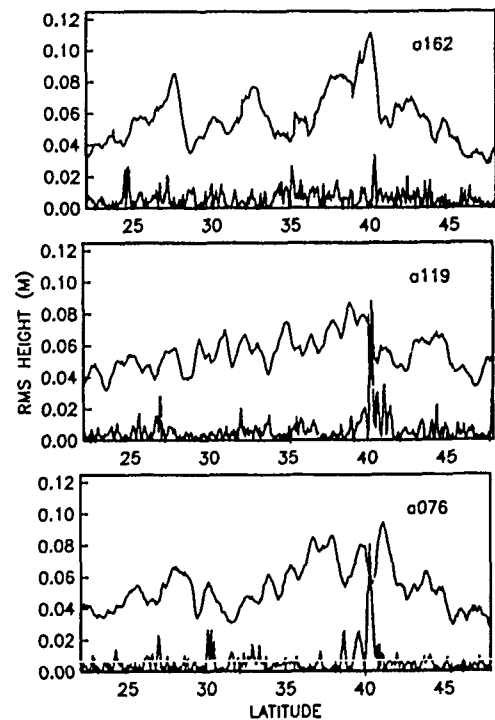


Fig. 6. Variability along the three ascending tracks nearest shore as function of latitude. Also shown is the rms longitude deviation from the exact repeat track, scaled by the along-track geoid slope.

Variability is enhanced in areas where meanders of the California Current have been observed previously. High variability off Baja California suggests the occurrence of similar meanders and filaments there. The band of low variability off Central California is a surprise. A strong offshore jet has been observed at that location in July 1934 (Chelton *et al.*, 1987) and July 1985 (Flament and Armi, 1989). Is this jet intermittent, so that this study is based on time series too short to be considered stationary? Or is this jet steadier than those off Northern California, leading to less variability of relative topography?

5. Propagation of features and temporal variations

Two-dimensional time-latitude correlation functions of $h(\lambda, t)$ are shown in Fig. 7 for the three ascending tracks near the coast. Along-track propagation of sea level anomalies at about 4 km/day poleward is evident for the two tracks nearest shore, but not further offshore. A similar northwestward propagation speed has been reported by Flament and Armi (1989), who observed the translation at ~ 4.7 cm/s of the root of an offshore jet near 36°N , using repeated hydrographic sections and sequences of satellite infrared images. However, there have also been reports of southward motion (e.g. Flament *et al.*, 1985).

Contours of relative topography as a function of λ and t are shown in Fig. 8 for the three ascending tracks near to the coast. Along the track nearest the coast (a162, 85 km from Point Arena), sea level drops sharply in mid-April and stays low throughout the spring and summer, i.e. throughout the upwelling season. Along the second offshore track (a119, 215 km from Point Arena), sea level is also low in summer and early fall, but the lowest levels are not seen until September. Along the third offshore track, (a076, 345 km from Point Arena), low sea levels are not seen until winter, and do not appear to be more persistent than high sea levels. These data, together with the hydrographic surveys, suggest that the lows in sea level are regions inshore of an upwelling front, which forms on the shelf in early spring and subsequently moves offshore.

Along-track variability, defined as $\Delta\lambda^{-1} \int_{\lambda}^{\lambda+\Delta\lambda} d\lambda h(\lambda, t)^2$,

was also computed for each track over the latitude band 30°N to 40°N , and is shown in Fig. 9 gray-coded as a function of time and of the longitude at which each track crosses 35°N . The 6-cm rms contour is outlined in white; dark areas correspond to larger values (up to 12-cm rms for black), light areas to smaller values. This plot is somewhat noisy, but nevertheless show a seasonal cycle. Along the track nearest the coast, along-track variability is smaller than 6 cm rms from late fall to late winter, and larger than 6 cm rms in spring and summer. The time at which along-track variability increases in spring occurs later further offshore, with a delay of ~ 0.5 day/km. The seasonal cycle decreases in amplitude towards offshore, and is not detected further than 129°W (~ 500 km from the coast). Removing for each track the mean sea level over the same latitude band does not modify significantly the contour plot, indicating that most of the along-track variability signal is due to the offshore jets at a scale of a few hundreds of kilometers and not to larger scale seasonal variations of sea level.

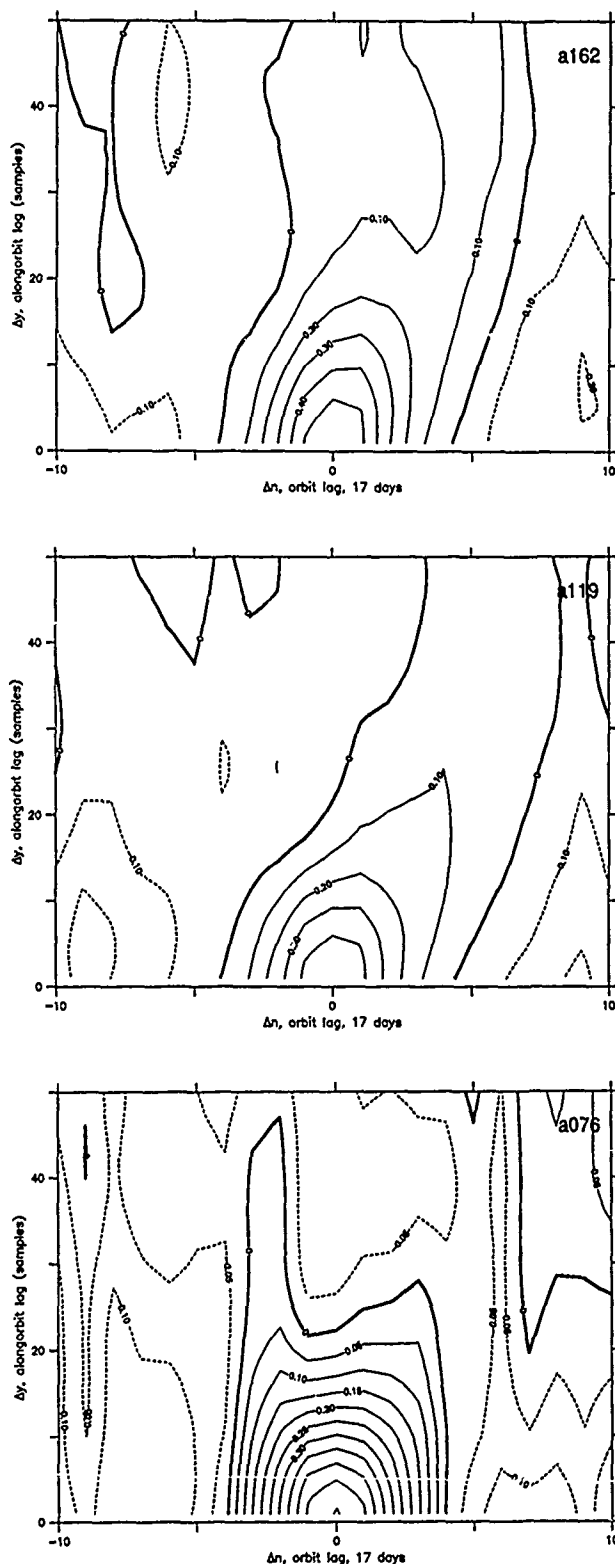


Fig. 7. Two-dimensional time-latitude correlation functions for tracks a162, a119 and a076. Lag units are in cycles (17 days) on the x axis and along-track samples (7.3 km) on the y axis.

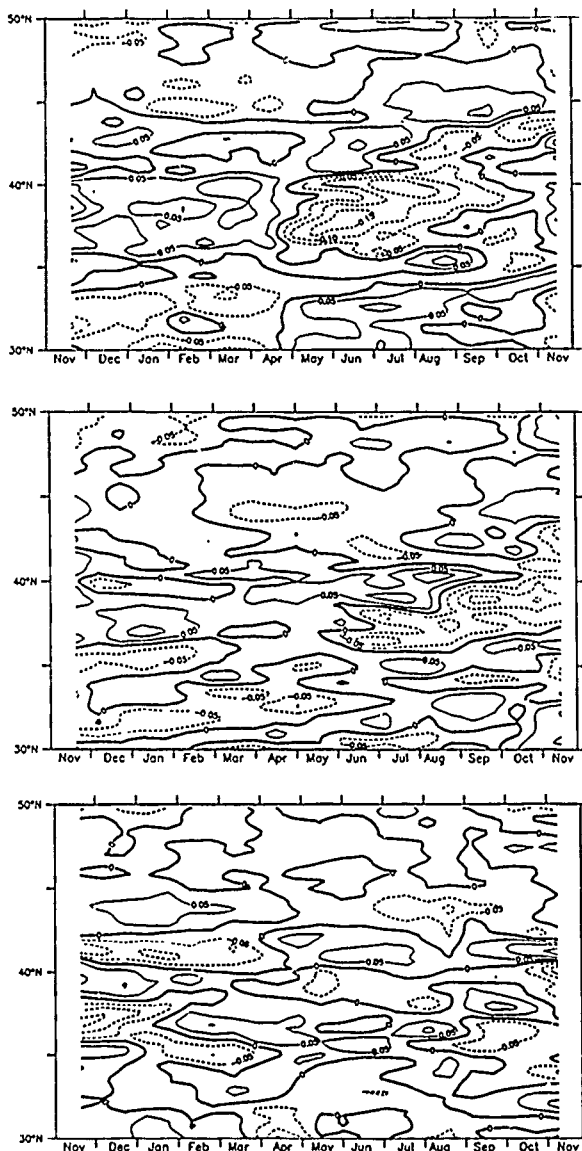


Fig. 8. Contours of relative topography as a function of time and latitude for tracks a162, a119 and a076. Negative levels are dashed.

6. Sources of errors

The principal sources of contamination are water vapor in the troposphere, cross-track geoid slopes and electro-magnetic bias. As will be shown below, none of them seem to be sufficient to affect the patterns presented above.

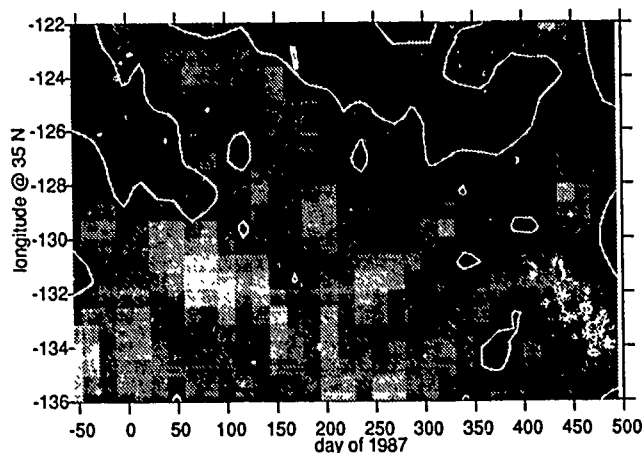


Fig. 9. Average along-track variability between 30°N and 40°N as a function of time and longitude at which each track crosses 35°N. The 6 cm rms contour is outlined. Darker areas correspond to larger values, lighter ones to smaller values.

The weakest correction is for microwave propagation delay due to water vapor in the troposphere. The correction provided on the tapes is derived from the 2.5°-resolution FNOC atmospheric model. It is of the order of 12 cm in this area, and varies by ± 6 cm, but on scales larger than the patterns of variability seen in Fig. 5. Fu (1983) presented power spectra of water vapor correction and surface topography, based on simultaneous SEASAT SMMR and altimeter data in low mesoscale energy areas, which showed a signal-to-correction ratio ranging from 12 dB at scales of 100 km to 4 dB at 1000 km.

Plots of rms longitude deviation from the exact repeat track, scaled by the along-track geoid slope, are shown with the plots of rms variability in Fig. 6 to provide a crude estimate of contamination by unknown small-scale cross-track geoid slopes, which can be large near seamounts and fracture zones. In particular, the Mendocino escarpment near 40°N contributes significantly to rms variability of the surface, and it is not clear whether the region of enhanced variability along 40°N is entirely real, or is in part an artifact due to cross-track deviations over the escarpment. Except for isolated seamounts, bottom topography is fairly regular further south and does not affect the patterns of variability.

Electro-magnetic bias is a wave-height dependent correction due to stronger reflection of the radar signal by the troughs than by the crests of asymmetric finite amplitude surface waves. It was not applied here. For GEOSAT, it is $\sim 1\%$ of significant wave height $H_{1/3}$ (Cheney *et al.*, 1988). Over the Eastern Pacific, $H_{1/3}$ measured by GEOSAT increases progressively from 2 m at 25°N to 4-5 m at 45°N, with little structure at scales of 1000 km or less. Most of the EM bias is thus removed with the parabolic orbit error.

7. Summary and conclusions

A mesoscale meander observed in the hydrography was also found in the altimeter topography. The measurements agreed to ~3-4 cm rms, showing that the GEOSAT altimeter is capable of resolving the mesoscale structure of the California Current.

The altimeter data showed that sea level variability is enhanced two to three-fold within 500 km from the coast of North America compared to the offshore background level. The strength of the variability varied also along the coast and peaked where strong meanders and westward jets have been found recurrently, suggesting that nearshore variability of relative topography can be attributed primarily to these meanders and jets.

Variability is modulated seasonally, suggesting that these jets are dynamically linked to the coastal upwelling cycle, and that the corresponding cold filaments observed in satellite images may not be simply the expression of colder nearshore water passively advected in a background field of offshore eddies.

The ~500 km offshore extension of high variability in sea level could be used as a definition of the Coastal Transition Zone. It is larger than the typical ~300 km extension of cold filaments seen in AVHRR images, indicating that the surface temperature expression of a filament disappears closer to shore than does the kinetic energy of the jet.

References

- Coastal Transition Zone Group, "The Coastal Transition Zone Program," *EOS, Trans. Am. Geophys. Union*, vol. 69, no. 27, p. 698ff (1988).
- Barth, J. A. and K. H. Brink, "Shipboard acoustic doppler profiler velocity observations near Point Conception: spring 1983," *J. Geophys. Res.*, vol. 92, pp. 3925-3943 (1987).
- Brink, K.H., "The near-surface dynamics of coastal upwelling," *Prog. Oceanog.*, vol. 12, pp. 223-257 (1983).
- Brink, K.H., "Coastal ocean physical processes," *Rev. Geophys.*, vol. 25, pp. 204-216 (1987).
- Chelton, D. B., R. L. Bernstein, A. Bratkovich, and P. M. Kosro, "The Central California Coastal Circulation Study," *EOS, Trans. Amer. Geophys. Union*, vol. 68, no. 1, p. 1ff (1987).
- Cheney, R.E., B.C. Douglas, and L. Miller, "Evaluation of GEOSAT altimeter data with application to tropical Pacific sea level variability," *J. Geophys. Res.* (1989). (in press)
- Cheney, R.E., J.G. Marsh, and B.D. Beckley, "Global mesoscale variability from collinear tracks of SEASAT altimeter data," *J. Geophys. Res.*, vol. 88, p. 4343 (1983).
- Flament, P. and L. Armi, "Observations of surface convergence, subduction and related finestructure," *J. Geophys. Res.* (1989). (submitted)
- Flament, P., L. Armi, and L. Washburn, "The evolving structure of an upwelling filament," *J. Geophys. Res.*, vol. 90, pp. 11,765-11,778 and 11,835-11,836 (1985).
- Fu, L.L., "On the wave number spectrum of oceanic mesoscale variability observed by the SEASAT altimeter," *J. Geophys. Res.*, vol. 88, pp. 4331-4342 (1983).
- Kosro, P.M. and A. Huyer, "CTD and velocity surveys of seaward jets off Northern California," *J. Geophys. Res.*, vol. 91, pp. 7680-7690 (1986).
- Rienecker, M., C.N.K. Mooers, D.E. Hagan, and A.R. Robinson, "A cool anomaly off Northern California: an investigation using IR imagery and in situ data," *J. Geophys. Res.*, vol. 90, pp. 4807-4818.
- Sailor, R.V. and A.R. LeSchack, "Preliminary determination of the GEOSAT altimeter noise spectrum," *Johns Hopkins APL Tech. Digest*, vol. 8, no. 2, pp. 182-183 (1987).
- Sailor, R.V. and A.R. LeSchack, "Preliminary determination of the GEOSAT altimeter noise spectrum," *Johns Hopkins APL Tech. Digest*, vol. 8, no. 2, pp. 182-183 (1987).
- Schramm, Richard E., Jane Fleischbein, Adriana Huyer, P. Michael Kosro, Tim Cowles, and Nan Dudek, "CTD Observations in the Coastal Transition Zone off Northern California, 9-18 June 1987," Ref. 88-3, College of Oceanography, Oregon State University, Corvallis, 1988b.
- Schramm, Richard E., Jane Fleischbein, Robert Marsh, Adriana Huyer, P. Michael Kosro, Tim Cowles, and Nan Dudek, "CTD Observations in the Coastal Transition Zone off Northern California, 18-27 May 1987," Ref. 88-3, College of Oceanography, Oregon State University, Corvallis, 1988a.

Acknowledgements. We would like to thank M. Caruso, S. Gille and Z. Sirkes, who assisted with the development of GEOSAT processing software, and D. Kelley for providing an outstanding plotting package. This work was partially supported by the Office of Naval Research through contracts N000014-87-K-007, N00014-87-K-0242 and N00014-86-K-0751, and grant N00014-89-J-1599.

Eddy scales resolved by the Geosat radar altimeter

James J. Bisagni

US Naval Underwater Systems Center
Newport, RI, USA

Abstract

Residual sea height anomalies (RSHA) were computed along 12 Geosat ground tracks from the Frontal Air-Sea Interaction Experiment (FASINEX) domain in the western North Atlantic during winter 1986. The RSHA were compared with residual sea surface temperature anomalies (RSSTA) extracted along the same tracks from infrared satellite images. RSHA and RSSTA ranged between ± 20 cm and ± 1 °C respectively. Numerical correlations showed that positive (negative) RSHA were correlated with positive (negative) RSSTA. Ensemble mean spectra of both RSHA and RSSTA exhibited a peak near $2.5 \times 10^{-3} \text{ km}^{-1}$, although spectral slopes between that scale and 10^{-2} km^{-1} appeared different. This suggests that the correlations existed primarily at a scale characteristic of mesoscale eddies; a scale well-documented from work completed in the same region during the Mid-Ocean Dynamics Experiment (MODE).

Satellite altimetry, AVHRR imagery, mesoscale eddies

Introduction

The classified 18 month Geosat Geodetic Mission (GM) coincided with the Frontal Air-Sea Interaction Experiment (FASINEX) conducted in the western North Atlantic during the first half of 1986. During the GM, Geosat was constrained to a non-repeating orbit, with the resulting residual sea heights being unavailable due to their classified nature. However, a preliminary unclassified data set, consisting of 3 months (January-March) of Geosat residual sea height anomalies (RSHA) was released for the FASINEX domain by the US Naval Oceanographic Office (NAVOCEANO) after the classified residual sea heights were differenced with the region's classified NAVOCEANO geoid. These unclassified data were then used in the following analysis.

The primary FASINEX objective was to determine the role of horizontal variability on air-sea interaction in the western North Atlantic subtropical convergence, located over the deep Hatteras Abyssal Plain. Upper ocean variability in this region manifests itself at the ocean's surface as a series of convoluted, but generally zonal sea surface temperature (SST) fronts, with individual

fronts of order 1-2 °C over a few km, (Voorhis, 1969). SST monitoring of the domain both prior to and during the experiment, using Advanced Very High Resolution Radiometer (AVHRR) satellite imagery, allowed both the locations and intensity of individual frontal features to be ascertained.

The FASINEX domain (Fig. 1) closely corresponded to that of the Mid-Ocean Dynamics Experiment (MODE) in which, the presence of mid-ocean mesoscale eddies has been well-documented, (Robinson, 1975; McWilliams, 1976). Furthermore, from additional analysis of MODE data, Voorhis et al. (1976) observed that SST appeared correlated with surface advection associated with the MODE mesoscale eddy field. Therefore, it was of interest to try and examine in a preliminary way, the relationship of mesoscale eddies present during FASINEX to the evolution of the SST field, using Geosat RSHA together with AVHRR satellite imagery.

Methods

Altimeter Data Processing

The Geosat RSHA corresponded to 12 (7 ascending + 5 descending) of the 27 passes released by NAVOCEANO for the January-March

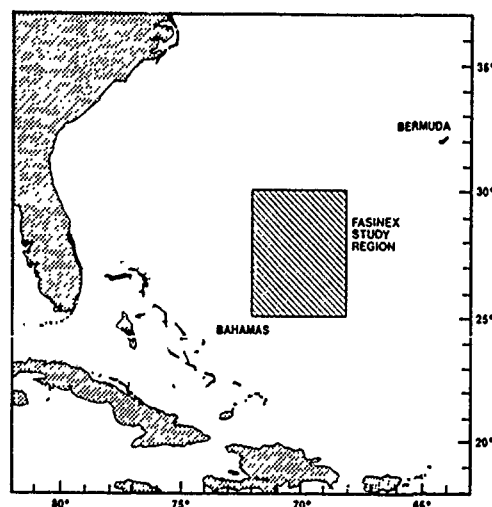


Figure 1. Location of the FASINEX region.

1986 time period. These 12 realizations (Fig. 2) represent all passes longer than 400 km, thus capable of resolving MODE-scale eddies with the characteristic diameter of 200 km. The data consisted of 2 Hz averages computed from the original 10 Hz data after removal of "bad" points, with successive samples separated by approximately 3.3 km. In addition to being differenced with the classified NAVOCEANO geoid, the data were also corrected for both ocean and earth tides, inverse-barometer effect, ionospheric and dry-tropospheric effects and surface waves. The data were not corrected for water vapor. However, radiosondes and surface meteorological observations collected during FASINEX indicated a temporal variation in the wet-tropospheric range correction of 15 cm, but with only minor spatial variability occurring up to length scales of nearly 400 km, (Bisagni, 1989). Each pass was de-trended using a linear model to remove orbit error and any other remaining long-wavelength trends. The data was low-pass filtered using a Gaussian filter with a half-power wavelength of 90 km.

The NAVOCEANO geoid is a "hybrid" surface, constructed from both satellite altimeter data (derived from the entire 18 month Geosat GM) and detailed gravity data collected from ships. The final surface was produced by filtering both data sets, emphasizing the altimeter data at long wavelengths and the gravity data at short wavelengths, (W. Rankin, pers. comm.). It is reasonable to assume that this geoid represents the best reference surface available for this region. This assumption is based on the types of data used to construct the surface, the region's low bathymetric relief and the good agreement shown between altimeter-derived mean sea surfaces and gravimetric geoids in the region, (Marsh et al., 1984).

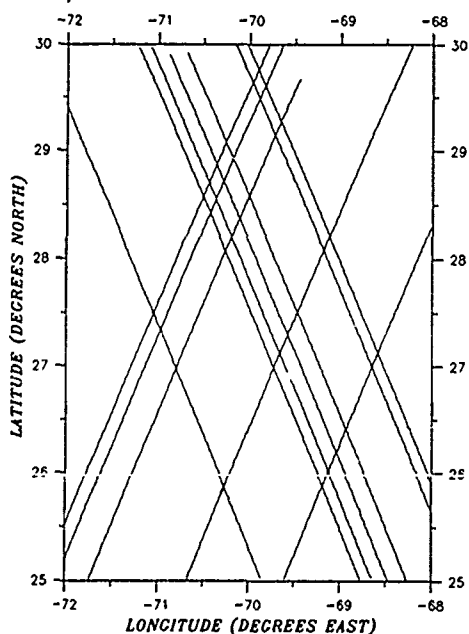


Figure 2. Locations of the 12 Geosat groundtracks.

AVHRR Image Processing

SST data derived from AVHRR images were obtained from channels 4 and 5 of the AVHRR/2 sensor aboard the NOAA-9 polar orbiter, using the NOAA/NESDIS (1982) algorithm. Data were then scan-angle corrected to compensate for atmospheric attenuation at large scan angles, (Cornillon et al., 1987). Sub-sampling of the data was accomplished by retaining the warmest pixel in each 4×4 pixel square. Images corresponding closely in time to the 12 Geosat passes described above were selected for further processing. Cloud areas present within these images were flagged and then enlarged slightly to eliminate cloud-edge effects. The groundtrack of each Geosat pass was then plotted on 1 or more candidate images. This procedure determined an "optimum image" with respect to minimizing the lag/lead time of the image relative to the day of the Geosat pass while maximizing the length of the cloud-free segments along the groundtrack.

Table 1
AVHRR Imagery Lag/Lead (Days)
Relative to the Date of Each Geosat Pass

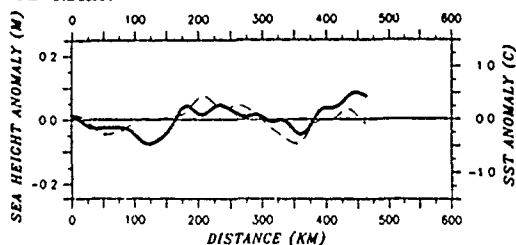
| Geosat Pass and Date | AVHRR Image(s)* and Date(s) | Lag/Lead +/- |
|-------------------------|--------------------------------|-----------------|
| A4237 1-03-86 | K86002 1-02-86 | -1 |
| D4259 1-04-86 | K86004 1-04-86 | 0 |
| D4302 1-07-86 | K86005 1-05-86 | -2 |
| | K86007 1-07-86 | 0 |
| D4345 1-10-86 | K86005 1-05-86 | -5 |
| | K86007 1-07-86 | -3 |
| A4567 1-26-86 | K86025 1-25-86 | -1 |
| | K86026 1-26-86 | 0 |
| A4897 2-18-86 | K86048 2-17-86 | -1 |
| | K86051 2-20-86 | +2 |
| D4919 2-19-86 | K86048 2-17-86 | -2 |
| | K86051 2-20-86 | +1 |
| A4940 2-21-86 | K86048 2-17-86 | -4 |
| | K86051 2-20-86 | -1 |
| A5184 3-10-86 | K86066 3-07-86 | -3 |
| | K86067 3-08-86 | -2 |
| A5227 3-13-86 | K86073 3-14-86 | +1 |
| | K86074 3-15-86 | +2 |
| | K86076 3-17-86 | +4 |
| | K86077 3-18-86 | +5 |
| D5249 3-15-86 | K86073 3-14-86 | -1 |
| | K86074 3-15-86 | 0 |
| | K86076 3-17-86 | +2 |
| | K86077 3-18-86 | +3 |
| A5270 3-16-86 | K86073 3-14-86 | -2 |
| | K86074 3-15-86 | -1 |
| | K86076 3-17-86 | +1 |
| | K86077 3-18-86 | +2 |

* Multiple images signify a composite

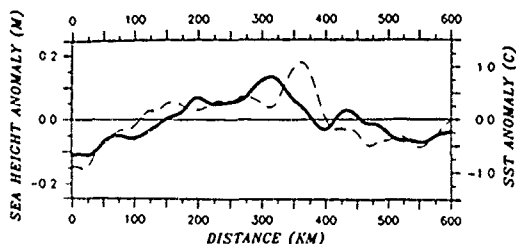
During cloudy periods when large portions of the SST images were obscured, "composite" images were constructed from (generally) 2 images which "bracketed" the date of the corresponding Geosat pass. In this process, both images were compared, 1 pixel at a time, with the result that only the "warmest" pixel from either image was retained in the composite. This algorithm assumed that the "cooler" of the 2 pixels was in some way "contaminated" by small clouds not detected earlier. Lag/lead times of all AVHRR data contained in the images relative to the days of the corresponding Geosat

passes are displayed in Table 1. The statistics show that 7 of the Geosat passes corresponded to AVHRR data having lag/lead values of ± 2 days or less, while only 3 of the passes had values of ± 4 or 5 days. Based on these statistics and the methods used, it is reasonable to assume that cloud-free AVHRR SST data, located along each of the 12 Geosat ground tracks, represent good estimates of the true along-track SST values. For analysis purposes, a single AVHRR image (non-composite or composite) was chosen and assigned to each of the 12 Geosat ground tracks.

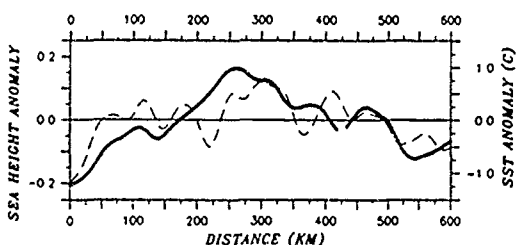
SST values were extracted from the corresponding AVHRR image at every point along each Geosat groundtrack. Resulting SST values were then de-trended and filtered using the same linear model and low-pass filter used to process the RSHA. The final result was a set of residual SST anomalies (RSSTA) which closely corresponded to each set of RSHA.



GEOSAT PASS A4237, 3 JAN 1986



GEOSAT PASS A4897, 18 FEB 1986



GEOSAT PASS A5227, 13 MAR 1986

Figure 3. Examples of corresponding RSHA (heavy line) and RSSTA (dashed) from 3 Geosat passes.

Results

RSHA from 3 of the Geosat tracks together with the corresponding RSSTA are shown in Figure 3. In this figure, RSHA and RSSTA range over values of about ± 20 cm and $\pm 1.0^\circ\text{C}$ respectively. Furthermore, there appears to be a correlation between these 2 relatively weak signals: positive RSHA

generally correspond to positive RSSTA and negative RSHA generally correspond to negative RSSTA, an inference confirmed through formal statistical analysis.

Cross-Correlation Analysis

Normalized cross-correlation coefficients were computed between RSHA and RSSTA for all 12 of the Geosat groundtracks at zero km spatial lag to examine the relationship between RSHA and RSSTA. The resulting values (R_0) for each of the 12 groundtracks and their statistical significance are given in Table 2. Of the 12 realizations, 10 exhibit positive correlations which are statistically significant at the 95% level, while the remaining 2 show positive but insignificant correlations.

Spectral Analysis

Along-track, one-dimensional wave number spectra were computed for the RSHA and RSSTA data sets corresponding to each Geosat groundtrack, (Figs. 4a, b). These spectra were computed using a standard fast Fourier transform (FFT), technique which treated each spatial series as a single, cosine-tapered segment for maximum spectral resolution. However, trade-offs for this maximum resolution included large error magnitudes. To reduce this error to an acceptable level, ensemble mean RSHA and RSSTA spectra were computed using all 12 individual sea height and SST spectra, (Figs. 5a, b). Results show that both ensemble mean spectra exhibit a broad peak near $2.5 \times 10^{-3} \text{ km}^{-1}$ corresponding to a 200 km half-wavelength. This value closely corresponds to the characteristic MODE-scale and therefore may indicate the presence of mesoscale (MODE) eddies within the FASINEX domain. Of equal interest, is the behavior of the slope of each spectrum at wavelengths shorter than the MODE-scale. Between the MODE-scale and wavelengths of 100 km, the RSHA spectrum exhibits an approximate k^{-3} slope, (k is wave number). In the same interval, the RSSTA spectrum reveals a noticeably different slope of k^{-2} .

Discussion

In the present study, the positive spatial correlations, together with the existence of the MODE-scale in both of the ensemble spectra, suggest that sea height and SST may be related at the MODE scale. Assuming that altimeter-derived sea heights are able to resolve baroclinic MODE-scale eddies, two different theories have been postulated regarding what the relationship should be between such eddies and their RSHA and SST. Extending earlier work done with Gulf Stream rings, Williams (1988) recently proposed that air-sea interaction may cause modification of mid-ocean eddies through deepening of the mixed layer after the summer solstice with the result that the eddy's deeper temperature structure is revealed in the SST field. Application of his model to a cyclonic eddy (a "low") in the northeast Atlantic suggested that the eddy's negative (cold) temperature anomaly might become entrained into the surface waters with a resulting decreased SST relative to surrounding waters. Thus the SST anomaly in this case

would be horizontally coherent with both the deeper eddy structure and its associated "low" in sea height.

In addition to the present study, other workers also report observations supporting this entrainment theory. Roden (1981) showed qualitative correlation between SST and dynamic height from the FRONTS 80 experiment undertaken during winter in the vicinity of the North Pacific subtropical front. Also from FRONTS 80, Van Woert (1982) showed significant positive correlations between mesoscale SST and subsurface temperatures to a depth of 800 m.

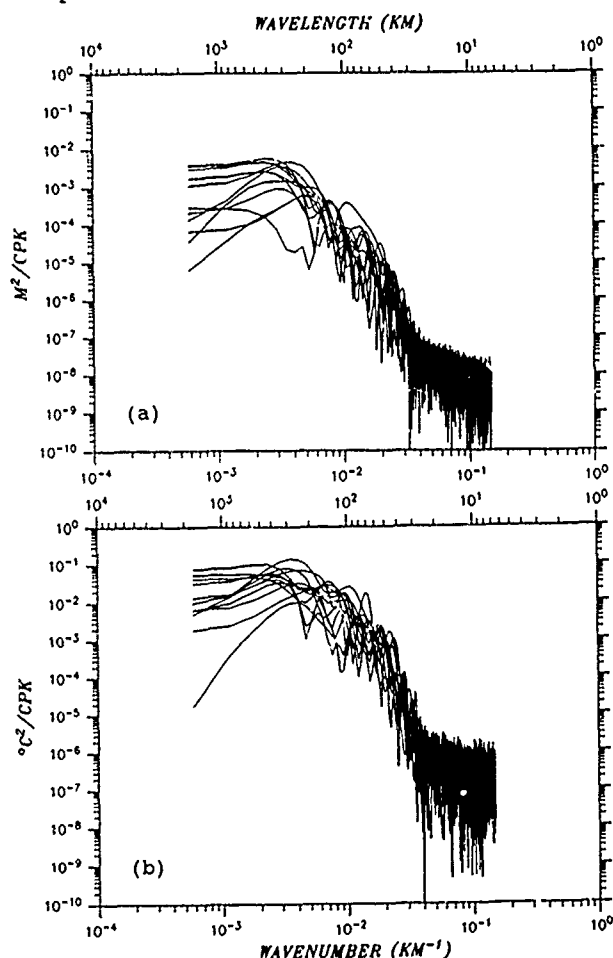


Figure 4. Composite, along-track wave number spectrum of: (a) RSHA, (b) RSSTA from each of 12 Geosat passes.

A second theory regarding the relation between sea height and SST resulted from the spring 1973 MODE work. During MODE, zonal and meridional SST features on scales of 40-400 km, seemed to result from advection caused by currents associated with the deep mesoscale eddy field as defined by hydrographic survey work (Voorhis et al., 1976). The features may have been the result of simple distortion of the mean meridional SST gradient into a series of alternating warm and cold intrusions located along the periphery of the eddies and separated by sharp fronts as

Table 2
Normalized Cross-Correlation Coefficients, R_0
At 0 km Lag Between 12 Sets of RSHA and RSSTA

| Pass | Date | Points Used | R_0 | 95% Signif. (?) |
|-------|---------|-------------|-------|-----------------|
| A4237 | 1-03-86 | 140 | 0.520 | YES |
| D4259 | 1-04-86 | 170 | 0.454 | YES |
| D4302 | 1-07-86 | 160 | 0.257 | YES |
| D4345 | 1-10-86 | 120 | 0.380 | YES |
| A4567 | 1-26-86 | 180 | 0.431 | YES |
| A4897 | 2-18-86 | 180 | 0.749 | YES |
| D4919 | 2-19-86 | 176 | 0.122 | NO |
| A4940 | 2-21-86 | 148 | 0.476 | YES |
| A5184 | 3-10-86 | 136 | 0.724 | YES |
| A5227 | 3-13-86 | 180 | 0.703 | YES |
| D5249 | 3-15-86 | 164 | 0.082 | NO |
| A5270 | 3-16-86 | 158 | 0.427 | YES |

reported by Fedorov (1983) from numerical models. Later work with the MODE data (Voorhis and Bruce, 1982) showed that in addition to advective SST variations, the primary cyclonic mesoscale eddy appeared

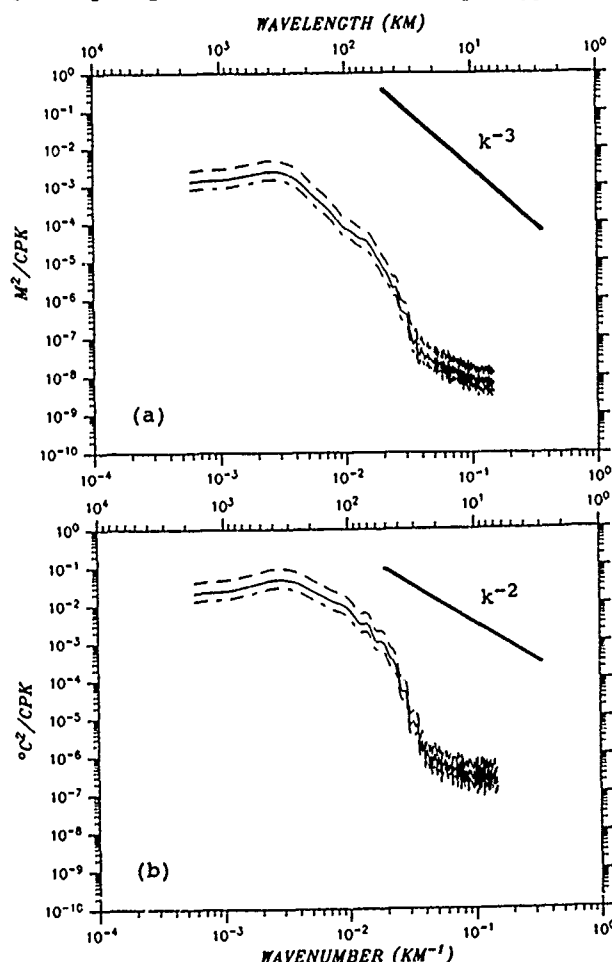


Figure 5. Along-track, ensemble mean wave number spectrum of: (a) RSHA, (b) RSSTA computed from all 12 Geosat passes. Also shown are the 95% confidence limits (dashed) and a reference slope.

centered over the coldest water in the region, in qualitative agreement with the above-described entrainment theory. Since SST structures resulting from advection would tend to show low coherence with the deep eddy field and sea height, Fedorov's advection theory may not be the only process responsible, as evidenced by the significant correlations noted in this study. Such structures, however, could be a rich source of noise, thereby reducing the correlations to the observed levels.

Comparison between the spectral slopes determined herein with those reported by other workers may also be useful in understanding the relationship between sea height and SST. However, it has been shown that caution must be used when interpreting any spectrum computed from a turbulent field, (Armi and Flament, 1985). This is due to the non-uniqueness of the turbulent field corresponding to any given spectrum, caused by the loss of phase information when the spectrum is computed. Acknowledging this caution, altimeter-derived sea height wave number spectra, computed in the western North Atlantic by Gordon and Baker (1980), Fu (1983) and Menard (1983) were examined. All show evidence of MODE-scale transients and nominal spectral slopes equalling k^{-2} , k^{-5} and k^{-3} respectively. The RSHA spectral slope of k^{-3} determined in the present study falls midway within their range and thus may represent a reasonable estimate of the mean spectral slope. Differences reported herein between RSHA and RSSTA spectral slopes may signify the presence of surface-intensified frontal features (and smaller scale processes) caused by mesoscale, eddy-induced advective structures. At these larger wave numbers, such phenomena may cause higher RSSTA energy (since they are detectable by AVHRR imagery) while having a minimal effect on RSHA energy. This is due to the small incremental effect they would have on sea height with respect to the much larger changes caused by mesoscale eddies. Future work with larger data sets from the region may be able to resolve this issue.

References

1. Armi, L. and P. Flament, 1985, Cautionary remarks on the spectral interpretation of turbulent flows. *J. Geophys. Res.* v. 90, p. 11779-11782.
2. Bisagni, J. J., 1989, Wet tropospheric range corrections for satellite altimeter-derived dynamic topographies in the western North Atlantic. *J. Geophys. Res.*, v. 94, 3247-3254.
3. Cornillon, P., C. Gilman, L. Stramma, O. Brown, R. Evans, and J. Brown, 1987, Processing and analysis of large volumes of satellite-derived thermal infrared data. *J. Geophys. Res.*, v. 92, p. 12993-13002.
4. Fedorov, N. K., 1983, The physical nature and structure of oceanic fronts. *Lect. Notes on Coast. and Estuar. Stud.*, 19, N. Demidenko, Translator, C. Garrett, Tech. Ed., Springer-Verlag, Berlin, 333 pp.
5. Fu, L.-L., 1983, On the wave number spectrum of oceanic mesoscale variability observed by the SEASAT altimeter. *J. Geophys. Res.*, v. 88, p. 4331-4341.
6. Gordon, A. L., and T. N. Baker, 1980, Ocean transients as observed by Geos 3 coincident orbits. *J. Geophys. Res.*, v. 85, p. 502-506.
7. Marsh, J. G., R. E. Cheney, J. J. McCarthy and T. V. Martin, 1984, Regional mean sea surfaces based on GEOS-3 and SEASAT altimeter data. *Mar. Geod.* v. 8, p. 385-402.
8. McWilliams, J. C., 1976, Maps from the mid-ocean dynamics experiment: Part I. Geostrophic Streamfunction. *J. Phys. Oceanog.*, v. 6, p. 810-827.
9. Menard, Y., 1983, Observations of eddy fields in the Northwest Atlantic and Northwest Pacific by SEASAT data. *J. Geophys. Res.*, v. 88., p. 1853-1866.
10. NOAA/NESDIS, 1982, Coefficients presented at the 32nd SST Research Panel Meeting, Nat. Environ. Sat., Data, Inf. Serv., Nat. Oceanic Atmos. Admin., Suitland, M.D., September 30, 1982.
11. Robinson, A. R., 1975, The variability of ocean currents. *Rev. Geophys. Space Physics*, v. 13, 598-601.
12. Roden, G. I., 1981, Mesoscale thermohaline, sound velocity and baroclinic flow structure of the Pacific subtropical front during the winter of 1980. *J. Phys. Oceanog.*, v. 11, p. 658-675.
13. Van Woert, M. L., 1982, The subtropical front: Satellite observations during FRONTS 80. *J. Geophys. Res.*, v. 87, p. 9523-9536.
14. Voorhis, A. D., 1969, The horizontal extent and persistence of thermal fronts in the Sargasso Sea. *Deep-Sea Res.*, v. 16 (suppl.), p. 331-337.
15. Voorhis, A. D., E. H. Schroeder and A. Leetmaa, 1976, The influence of deep mesoscale eddies on sea surface temperature in the North Atlantic subtropical convergence. *J. Phys. Oceanog.*, v. 6, p. 953-961.
16. Voorhis, A. D., and J. C. Bruce, 1982, Small-scale surface stirring and frontogenesis in the subtropical convergence of the western North Atlantic. *J. Mar. Res.*, v. 40 (suppl.), p. 801-822.
17. Williams, R. G., 1988, Modification of ocean eddies by air-sea interaction. *J. Geophys. Res.*, v. 93, p. 15523-15533.

ANALYSIS OF THE FETCH-RELATED BIAS IN ALTIMETER WIND SPEED MEASUREMENTS

Roman E. Glazman and Stuart H. Pilorz

Jet Propulsion Laboratory. California Institute of Technology.

4800 Oak Grove Dr., Pasadena, CA 91109

ABSTRACT

The present interpretation of the radar cross section σ^0 measured by satellite altimeters implies that the rms wave slope γ is controlled solely by the local wind. However, parameters of wave spectra, including the exponent in the power law for the equilibrium range, depend on sea maturity. The latter is characterized by the non-dimensional fetch, $z = gX/U^2$. Consequently, γ and σ^0 are controlled both by U and by the wind fetch X . One year worth of GEOSAT data are used jointly with in situ wind and wave observations to assess the fetch-related error trend in altimeter wind speeds. This trend results in overestimated winds in the regions and seasons characterized by a high z , and vice versa. A procedure for wind speed retrieval based on processing σ^0 jointly with the significant wave height information contained in the altimeter wave forms is proposed.

1. THEORETICAL BACKGROUND

In our recent studies of the accuracy of wind speed measurements by the Seasat scatterometer [5,6], an error trend correlated with various measures of sea maturity was found. An explanation of the trend was suggested based on a model of the radar cross section comprised of two components, the traditional Bragg-scattering component σ_B^0 and the spike component σ_s^0 accounting for the contribution of the intermittently occurring extremely steep wavelets. The relative importance of the latter component was found to grow as the degree of wave development increases. This led us to conclude that the degree of wave development, rather than the wind speed alone, plays a certain role in determining basic statistics of the rms slope for gravity-range waves. The equilibrium range of wave spectra, in our view, is not confined to one of the commonly discussed forms, k^{-4} (the classical Phillips law) or $k^{-7/2}$ (the Zakharov-Filonenko spectrum). An extensive survey of both theoretical and experimental literature led us to accept a more general form for the equilibrium range [6,7,8] based on an incomplete-similarity theory due to Barenblatt and Leykin [1]:

$$F(k, \psi) = \beta (U^2/g)^{2\mu} k^{-1+2\mu} \Upsilon_k(\psi) \quad (1)$$

where $\beta = \beta(z)$ and $\mu = \mu(z)$; U is the mean wind speed at a height, say, 19.5 m and z may be interpreted as, for example, the non-dimensional wind fetch. Υ is the angular spread function. For open ocean conditions, parameter μ varies from about zero to 1/3, and according to some observations, e.g. [10], it can exceed 1/3. It has a simple interpretation as the fractal co-dimension of a small surface patch [8].

The corresponding wave slope variance, found by integrating equation (1) multiplied by k^2 , is a function of both the wind speed and the degree of wave development [7,8]. However, such integration is meaningless unless the wave spectrum has a rapidly decaying high-frequency tail, commencing at the wavenumber $2\pi/h$ corresponding to certain, intrinsic microscale h . The existence of the microscale is dictated by the hydrodynamic instability of excessively steep gravity wavelets. Based on this argument, the microscale has been estimated to lie between 1/2 and 1 m [4,7,8]. Although a specific choice of an averaging function realizing the spectral cutoff does not crucially affect the resulting value of the wave slope, we believe that the smooth spectral roll-off of the form $\exp[-(kh)^2]$ represents the best low-pass filter [8]. The slope variance is then given by

$$\gamma^2 \approx \int_0^\infty F(k) \exp[-(kh)^2] k dk \approx \frac{\beta(z)}{2} \left[\Gamma(\mu) - \frac{\delta^{2\mu}}{\mu} \right] \left(\frac{U^2}{gh} \right)^{2\mu} \quad (2)$$

where $\Gamma(\mu)$ is the gamma function, k_0 is the spectrum peak wavenumber, and $\delta = k_0 h$. Equation (2) implies that the small-scale roughness, which is not correctly represented by the form (1) anyway, is smoothed out. With respect to electro-magnetic scattering, the small-scale roughness with wavelengths under h determines the coefficient $V(\theta)$ for local coherent reflection at the faces of large-scale waves which pertain to the equilibrium range (1). θ is the incidence angle.

This leads to a simple model for near-nadir backscatter

$$\sigma^0 = \frac{|V(\theta)|^2 \exp[-(\tan\theta/\gamma)^2]}{\cos^4\theta \gamma^2} \quad (3)$$

in which $V(\theta)$ is a (presumably, rather weak) function of the local instantaneous wind speed and γ^2 is controlled both by the mean wind speed and by the degree of wave development. Evidently, as this degree increases (hence, μ in (1) grows), the wind dependence of the radar cross section becomes ever stronger.

One goal of the present work is to assess the importance of the x -dependent bias in altimeter wind measurements and to make a step towards correcting this bias. Another reason for this study is a potential capability of a satellite altimeter to indirectly measure the wave development. The reader is referred to the recent works [7,8] for details of the theory, and to [9] for details of the experimental approach.

2. EXPERIMENTAL RESULTS

Geosat altimeter data were colocated with twenty NDBC buoys conducting autonomous observations in various regions of the World Ocean. The colocation procedure was similar to that employed by Dobson et al. [3]. However, we included into our analysis only the cases of equilibrium sea states for which the theory presented in the previous section is relevant. To this end, we browsed through wave spectra

reported by the buoys for each individual case and selected the cases in which the spectra had a well expressed equilibrium range and a single peak. Using the significant wave height $H_{1/3}$, wave period T and wind speed U data from the buoys, we estimated the generalized wind fetch X and the wind-to-wave speed ratio ξ :

$$X = 3.4 \cdot 10^5 \frac{gH_{1/3}^2}{U^2} \quad (4)$$

$$\xi = \frac{C_0}{U} \approx \frac{gT}{2\pi U} \quad (5)$$

The altimeter wind speeds were estimated using the Brown "smoothed" algorithm [2], and the error

$$\epsilon = U_s - U_B \quad (6)$$

was plotted versus ξ , Figure 1. The coefficient of correlation of ϵ with ξ was found to be 0.51. The measure ξ given by (5) is based on the dispersion relationship for deep-water gravity waves. In [9] this measure was estimated differently: firstly we calculated the generalized non-dimensional fetch:

$$z = \frac{gX^2}{U^2} \quad (7)$$

using (4) to estimate X , and then we estimated the wave-to-wind speed ratio using the approximation derived earlier [5,6]:

$$\frac{C_0}{U} \approx 5.6 \cdot 10^{-2} z^{0.3} \quad (8)$$

However, comparing Figure 1 to the corresponding plot in [9] one finds that the results are very similar.

Similar analysis was carried out for the significant wave height measurements by the altimeter. No trend related to any measure of wave development was found in the wave height errors.

3. APPROACH TO THE ALGORITHM DEVELOPMENT

To eliminate the trend from the wind speed measurements, one has to know the actual degree of wave development. Since the altimeter provides information on both the wind speed and the wave height, a possible approach consists in using the radar cross section along with the wave height information. In order to realize this approach, let us investigate the actual dependence of the backscattered power on the wind speed and wind fetch. To this purpose, we represent the radar cross section in the form:

$$\sigma_{dB}^0 = P(X) + Q(X) \log U \quad (9)$$

where the coefficients P and Q are sought as functions of the generalized wind fetch, (4). Using the buoy data on wind speed and wave height, we estimated X for all colocated observations and broke the entire data set into eleven overlapping gradations of X . Curve (9) was fitted to data points in each gradation, as illustrated in Fig. 2, which resulted in a set of values $P(X_i)$ and $Q(X_i)$ for each i th gradations. In Figure 3 we show the corresponding dependences.

Obviously, these empirical dependences are sufficiently smooth to grant their representation in a functional (for instance, quadratic or cubic polynomial) form. Functions $P(X)$ and $Q(X)$ would allow one to solve the inverse problem: given the significant wave height $H_{1/3}$ and the radar cross section σ_{dB}^0 , reported by the altimeter, determine the wind speed and fetch. Two different approaches are possible.

1. The iterative approach.

Due to the fact that the radar cross section is much more sensitive to the wind speed than to the fetch, one way of solving the problem is envisioned as an iterative process. Specifically, in the zero-order approximation, the wind speed can be estimated using one of the traditional algorithms, such as the Brown smooth model function. The resulting wind and the altimeter wave height $H_{1/3}$ can then be used for a zero-order estimate of the generalized fetch (4). Based on this X , one finds the zero-order approximation for $P(X)$ and $Q(X)$ and solves (9) for U . This wind represents the first-order approximation, which can be iterated by repeating the process to obtain higher-order estimates of both wind speed and wind fetch.

2. Look-up tables.

Based on (9), the look-up tables to determine the wind speed U as a function of σ_{dB}^0 and $H_{1/3}$ can be calculated.

Our work on the development and verification of an unbiased geophysical model function will be reported in a separate publication, for this effort requires a more complete data set than that used in the present work.

ACKNOWLEDGEMENT. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautic and Space Administration.

REFERENCES

1. Barenblatt, G.I. and I.A. Leykin, 1981. On the self-similar spectra of wind waves in the high-frequency range. *Izvestiya, Atmospheric and Oceanic Physics*. (English Translation), 17(1), 35-41
2. Brown, G.S., H.R. Stanley, and N.A. Roy. 1981. The wind speed measurement capability of spaceborne radar altimeters. *IEEE J. Oceanic Eng.*, OE-6(2), 59-63
3. Dobson, E.B., F.M. Monaldo, J. Goldhirsh and J. Wilkerson, 1987. Validation of Geosat altimeter-derived wind speeds and significant wave heights using buoy data. *J. Geophys. Res.*, 92(C10), 10719-10731
4. Glazman, R.E., 1986. Statistical characterization of sea surface geometry for a wave slope field discontinuous in the mean square. *J. Geophys. Res.*, 91(C5), 6620-6641
5. Glazman, R.E., 1987. Wind-fetch dependence of Seasat scatterometer measurements. *Int. J. Rem. Sensing*, 8(11), 1641-1647
6. Glazman, R.E., C.G. Pihos and J. Ip. 1988. Scatterometer wind-speed bias induced by the large-scale component of the wave field. *J. Geophys. Res.*, 93(C2), 1317-1328
7. Glazman, R.F. Near-nadir backscatter from a well-developed sea. *IEEE Trans. Geosci. and Rem. Sensing*. Submitted: May, 1988.

8. Glazman, R.E. and P. Weichman, 1989. Statistical geometry of a small surface patch. *J. Geophys. Res., Oceans*. In press.

9. Glazman, R.E. and S.H. Pilorz, 1989. Effects of sea maturity on satellite altimeter measurements. Submitted to: *J. Geophys. Res., Oceans*. April, 1989.

10. Grose, P.L., K.L. Warsh and M. Garstang, 1972. Dispersion relations and wave shapes. *J. Geophys. Res.*, 3902-3906

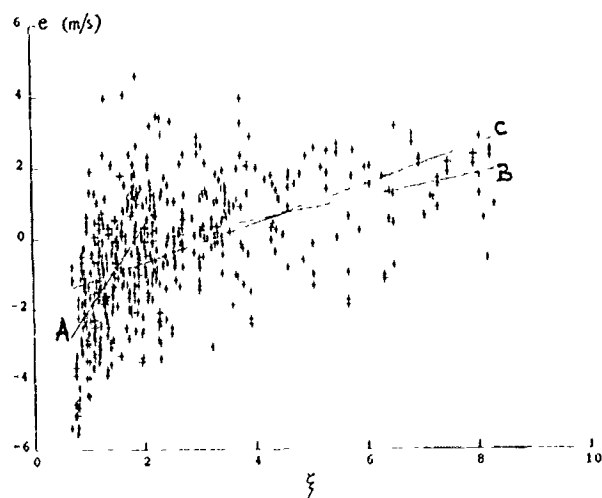


Figure 1. The absolute error of the altimeter measurements of the mean wind speed (6), versus the wave development measure ξ obtained from buoy data using (5). Straight line A approximates the trend on the interval $0 < \xi \leq 2$. Line B corresponds to $\xi > 2$, and line C represents the linear approximation for the entire range of ξ . The error trend for each segment is characterized by coefficients a and b : $e = a + b\xi$. These are found by linear regression: A: $a \approx -4.36$, $b \approx 2.33$; B: $a \approx -0.75$, $b \approx 0.32$; C: $a \approx -1.74$, $b \approx 0.55$.

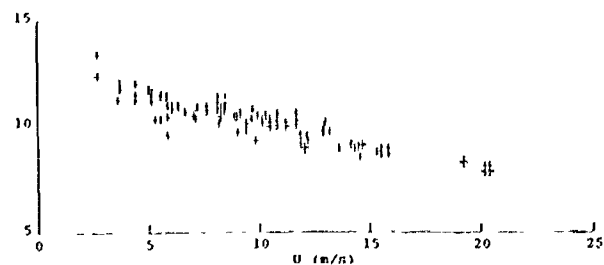


Figure 2. Wind-speed dependence of the radar cross section for one of the gradations of the generalized wind fetch, X , defined by (4). The dashed line represents the empirical fit described by (9).

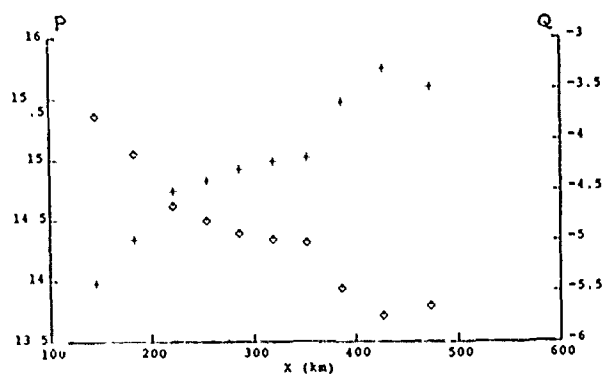


Figure 3. Fetch dependence of P (crosses) and Q (diamonds), (9), as based on empirical fits illustrated in Fig. 2.

GLOBAL ALTIMETER MEASUREMENTS OF EXTREMES OF
WIND SPEEDS AND WAVEHEIGHTS AND THEIR IMPACT ON FORECASTING

Ella B. Dobson
The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20707-6099
Telephone: (301) 953-5000, ext. 8645
FAX: 301-953-1093

Meteorologists have long known that weather patterns are cyclical in nature and in many cases these tracks can be accurately forecast within statistical accuracy. There are, however, many cases when accurate forecasting is not possible. More investigation of the spatial and temporal scales of intense weather patterns may lead to a better forecast method or at least augment the present systems.

This paper presents a study of extreme wind speeds and wave heights associated with intense weather patterns. Measurements of these parameters by the Geosat altimeter over a three year period are analyzed to determine spatial and temporal scales. Scales and movement in the southern oceans are compared to northern patterns. A discussion is presented on the potential usefulness of these statistics on forecasting. For example, in a region where Geosat data reveals a 80 percent probability of storm formation within a certain area, this region could be routinely monitored using the altimeter data. The question arises whether this would be valuable to the meteorological community.

SATELLITE ALTIMETRY: UTILIZATION FOR RESOURCE EXPLORATION

D. Dellkaraoglou (1), J.A.R. Blais (2), E.R. Kanasevich (3), D. Agouridis (4)

(1) Canada Centre for Surveying, Geodetic Survey Division, 615 Booth Street, Ottawa, Ontario K1A 0E9

(2) Department of Surveying Engineering, University of Calgary, 2500 University Drive, Calgary, Alberta T2N 1N4

(3) Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1

(4) Geoscience Integrated Inc., SUITE 171, 6R15-8 Street N.E., Calgary, Alberta T2E 7H7

ABSTRACT

The current surge of technological change in the fields of satellite positioning and remote sensing, especially in view of the coordinated sequence of specialized satellite missions leading to the Earth Observing System (Eos) platforms in the mid-1990's and beyond, emphasizes the need to develop modern, quantitative, rigorous and integrated methodologies which will make use of those missions' unique capabilities. Today spaceborne altimeters provide height observations to ground, ice sheet and oceanic surfaces with precision of centimeters. These data can be directly related to measurable gravity changes and be used for the mapping of density contrasts of the upper crust and for detecting geological features with distinctive signatures in either their topography, gravity field or spectral reflectance.

In recent work we have been investigating the potential utilization of satellite altimetry in the area of petroleum exploration. In this paper we present a brief outline of the state-of-the-art in the instrumentation and specialized treatment of the data required in order to extract such information out of the information content of the altimetry measurements. The methodologies involved in extracting significant information in the high frequency component of existing satellite altimetry data are summarized in an attempt to illustrate how these procedures bring forth a completely new source of geophysical information that can be used in an integrated manner with the results from other prospecting tools in the endeavour to acquire as much geological understanding as possible for features of economic significance.

1. INTRODUCTION

Satellite altimetry is a technique that has been with us for more than a decade. GEOS-3 and SEASAT, the first truly useful altimetric satellites, collected data in the period 1975-1978 leading to a great deal of advancements, especially in the areas of oceanography and the knowledge of the gravity field. These achievements are summarized in a number of dedicated issues of the *Journal of Geophysical Research* (AGU, 1979; AGU, 1982; AGU, 1983).

The latest altimetric mission, GEOSAT, is in progress since March 1985. To date over 2 years worth of GEOSAT altimetry data is available to civilian scientists. This is the data set released following the November 1986 de-classification of the mission by the U.S. Navy. The preliminary assessment of GEOSAT results indicates that this latest data set is of very high quality approaching a precision of 3 cm (McDonathy and Kilgus, 1987) and hence likely to yield exciting new results. In the near future, the *Ocean Topography Experiment (TOPEX)* satellite, the *Earth Resource Satellite (ERS-1)*, the *Geoscience Laser Ranging (and*

Altimetry) System (GLRS) onboard one of the *Earth Observing System (Eos)* polar platforms, it is planned to include satellite altimetry instrumentation fundamental to the Eos goals and objectives.

Satellite radar altimetry has to date been extensively evaluated at longer wavelengths (> 300 km) for gaining improved insight, for instance, into:

- the global characteristics of the geopotential
- the geophysical and geological interpretation of gravity across oceanic ridges
- the thickness and rigidity of the crust
- isostatic mechanisms

and other analyses related to geophysics and geology. However, the direct applicability of spaceborne altimeters for locating earth resources has been by far a subject of speculation with little comprehensive treatment found in the literature. This is mainly due to the fact that for the purpose of locating earth resources, analysis of gravitational wavelengths of a few tens of kilometers or less must be undertaken. Up until recently, these short wavelengths have not been extensively evaluated from altimetry data due to their proximity to the threshold of resolution of the altimeter instrumentation itself and the specialized treatment of the data needed to extract this type of information out of the information content of the altimetry measurements. The latter is spanning the entire spectrum of the gravity field signal up to a resolution limited by the altimeter's accuracy, surface footprint and data density. The sea surface heights measured directly by spaceborne altimeters also contain information of non-gravitational nature such as sea state, eddies, ocean currents and tides which need to be eliminated in order to arrive at a static and time independent sea surface. Essentially the shape that the sea surface would have, if it were completely at rest and not disturbed by waves, tides, currents etc., would conform to the equipotential surface of the earth's gravity field known as the *geoid*.

2. SATELLITE ALTIMETRY AS GEOPHYSICAL TOOL

The usefulness of altimetric measurements for geophysical studies has been reviewed by the U.S. National Research Council (1979); in their findings is pointed out that such measurements can make major contributions to geophysical problems. Because the global dynamic sea-surface topography shows up as a signal with an amplitude of 2 m at the most, superimposed on geoid undulations of up to about 100 m in amplitude (relative to a reference ellipsoid) extending in wavelengths of a few to thousand of kilometers, it is clear that satellite altimeters observe

primarily the gravitational field and only secondarily other geophysical processes such as tides and ocean currents.

Short wavelength anomalies in the geoid arise from shallow variations in the distribution of mass in the earth's crust which have been shown to correlate well with the topography of the ocean floor and of shallow crustal structures. This observed response to topography contains information about the compensation mechanisms as well as the relative densities of the loads (e.g. seamounts) and underlying mantle. Satellite altimeters like those flown on SEASAT or GEOSAT, and more so those planned for TOPEX, ERS-1 and GLRS, can resolve short wavelength geoidal features with an accuracy of a few cm and with a resolution of some tens of km. Hence they can be used to observe the geoid effects of features such as seamounts, fracture zones, ridges, trenches and bathymetric swells associated with hot spots. Conversely, altimetry-derived estimations made for sea floor topography should further contribute to understanding the contributions of the near sub-surface, thus allowing the study in detail of that additional high-frequency information found in existing altimetry data and the results of correlating the produced geoidal maps to offshore areas of known hydrocarbon accumulations.

Salt domes are of particular interest and therefore of high economic importance in the exploration of oil because they deform the surrounding sediments yielding a place for oil to accumulate. This premise is supported by the fact that a large fraction of the existing off-shore oil fields are located in regions containing salt dome structures. The latter have vertical juxtaposition of material having strong density contrasts, i.e. the salt (halite) vis-à-vis the surrounding material (sediments). The lateral density variation is particularly visible from an analysis of the resulting changes in the gravity field measured in the vicinity of these domes. Unfortunately, salt domes are a geological formation existing in limited regions over the earth's surface. Their formation requires tectonically stable regions along the margin of the continental plates or in areas which were covered by large, shallow inland seas. Such areas are further restricted to those regions having had an environment conducive to the formation of oil bearing stratum, i.e. having been the site of accumulation of the remains of prolific marine life during the mesozoic, or (more rarely) in earlier periods. Despite such areal restrictions, however, the total remaining regions of interest still vastly exceed the capabilities of exploration using standard surveying techniques (e.g. marine gravimetry). In this regard, while considerable information already exists about the geological structure of the offshore continental margins, this information is incomplete and insufficient in itself to completely define the scope of oil exploration investigations. Yet the volume and quality of satellite altimetry information is rapidly increasing and the cost effectiveness of such remote sensing capability is such that areas lacking in-depth geological information are worth evaluating nevertheless, since the location of geoidal signatures of interest would provide important information for detailed in situ investigations.

3. GRAVITATIONAL SIGNAL ANALYSIS PRINCIPLES

In order to explore the satellite altimetry capabilities for detecting geophysically interesting geological features, there are two major aspects of analysis that are necessary:

(a) *the development of appropriate processing techniques to eliminate various errors and measurement outliers from the satellite altimetry data;*

(b) *the construction of a model of the geological signatures to be located in the altimetry data in conjunction with a statistical model of altimetry data in the absence of a given geological signature (be it a salt dome, a seamount, etc.).*

The following brief discussion outlines the major considerations involved in each of these aspects, particularly with

regard to the possibilities and limitations of the current processing techniques.

Satellite Altimetry Data Reduction

Satellite altimetry data is contaminated by many errors. The measured height of the altimeter above the instantaneous sea level at the instant of the measurement contains unmodelled instrumental and propagation errors, as well as the effect of waves within the footprint area of the radar (which is typically of the order of a few kilometers for the SEASAT- or GEOSAT-type altimeters, and of the order of a few hundred metres for the GLRS laser altimeter). In practice, ice floats or small islands also contaminate further the radar reflections. The data must then be processed through a set of smoothing and editing algorithms and filtered to remove errors caused by atmospheric effects, instrumental and known sea-state dynamic effects and other types of "noise" of wavelengths foreign to the task of undulation signature detection (Berube et al., 1988). For instance, temporal effects in the data may be investigated by comparing spatially collinear or repeat altimetric profiles and by constructing differenced time series, i.e. point-by-point differences of sea surface heights collected over several such passes. These differenced series will consist mostly of noise, since the time invariant part of the signal of interest, i.e. the geoid, nearly cancels out when the difference profile is computed from passes separated by less than 2 km. Hence it is possible that features which are seen to vary in time can be removed (or greatly reduced) by averaging the coincident data sets or removing linear trends from the altimetry profiles. Indeed, this approach has been shown to be very effective for providing sea surface measurements of high accuracy and resolution without the need for precise geoid or orbital information.

Of quite different nature, long-wavelength errors in the altimetry data are due mostly to radial orbit errors. Precision orbit determination for the altimetry satellites has been limited primarily by inadequate global tracking, errors in the various dynamic models and in particular in the Earth's gravity model. These errors reveal their presence as differences between ascending and descending passes at their respective crossover points. Cross-over minimization techniques are commonly applied to compute mean sea surfaces from satellite altimetry. This type of crossover adjustments is an essential element of the data analysis since this type of approach provides the opportunity to obtain a highly self-consistent data set and to exercise statistical selectivity, thereby reducing the adverse effects of orbit misalignment in the processing of the altimetry data. In the near future, it is expected that because of further anticipated improvements in the earth gravity field and other dynamic models until 1990 and the use of GPS tracking concepts involving a GPS receiver on the spacecraft and 10-15 dedicated tracking stations on earth (Young et al., 1988), the radial component of the orbits of future altimetry satellites like TOPEX may be computed with an accuracy of 20 cm or better.

Geoid Signatures of Oil-related Geological Structures

Large lateral density variations as those found between the geological structures of interest to oil exploration and the surrounding material result in differential changes in the local gravitational potential given by

$$\delta V(P) = k \iiint_V (\delta \rho(Q) / d_{PQ}) dV \quad (1)$$

where the integration is carried out over the volume V of the dome, $\delta \rho(Q)$ is the density contrast at Q , d_{PQ} is the distance between P and Q , and k is the gravitational constant. The gravitational signal in the corresponding gravity anomalies Δg 's is strongly related to the variations in the structure, depth and dimensions of the geological structure. The geoid signal revealed by satellite altimeters may be used instead of gravity for

important aspects of these studies. This is because the Fourier transform F_g of the vertical component of the gravity field is directly related to the Fourier transform of the F_N of the geoid undulation by the simple relationship

$$F_N = F_g / (G |k|) \quad (2)$$

where G is the mean value of gravity (9.8 m/sec^2) and $k = 2\pi/\lambda$ denotes the wavenumber associated with undulations of wavelength λ . In simple terms, the spectral relationship between gravity anomalies Δg and the geoidal heights N is given by

$$N_n = (R/G) [\Delta g_n / (n-1)] \quad (3)$$

where Δg_n and N_n are the n -th degree harmonic of the gravity anomalies and geoid undulations respectively, and R is the mean radius of the earth. From (3) is evident that the geoidal heights have the same spectral content as do gravity anomalies, except that the amplitude of the geoidal heights as scaled by a factor $(n-1)^{-1}$ which results in a reduction in amplitude with increasing degree. To date the distribution of ground tracks from GEOS-3, SEASAT and GEOSAT has achieved a sufficient density in most parts of the world oceans, making a relatively complete survey of salt domes and related geological features possible. This raises the possibility that satellite altimetry data should, in their own right, be used to map changes in the potential surfaces when the high-frequency spectrum of the geoid is sought as in the case of hydrocarbon prospecting applications.

Detection of specific geological features to be sought in the altimetry data requires that the spectral signatures of a large number of possible configurations of these features be known. That is, the sought undulation signature after regularization-i.e. after the standard reductions made to altimetry data to remove the effects of anticipated (modelled) gravitational signals due to topography (bathymetry), isostatic compensation, and reference gravity field variations-must be known either from prior experience or be adequately modelled. A similar approach was applied by Schwank and Lazarewicz (1982) and White et al. (1983) for the detection of seamounts and for subsequently inferring about the underlying compensation mechanisms. However, it should be realized that such detection is much simpler than relating the results to the physical properties of the signatures-which is the main purpose of using such a technique as a geophysical tool. This is because if, for instance, salt domes signatures are to be sought in the data, a family of all possible salt dome gravitational signal signatures must be defined first as a function of:

- the depth of the dome below mean sea level (0 to 10 km),
- the width of the dome (1 to 40 km),
- the cross over depth of the surrounding sediments (500 to 2000 m),
- the composition (density of 2.4 to 2.9 g/cm^3) and thickness (0.1 to 2 km) of the cap rock, and
- the density of the surrounding material (1.7 to 2.9 g/cm^3).

This is usually the result of a standard parametric study which presents in both the frequency and amplitude dimensions the anticipated magnitude of various types of signatures which must be isolated (detected) in the satellite altimetry profiles.

Definition of Spectral Characteristics of Altimetry data

In addition to the models of undulation signatures associated with specific geological signatures such as salt domes, their detection in altimetry data requires a spectral model of the altimetry system characteristics. This requirement necessitates from two reasons: (i) different altimetry system characteristics are mapped onto the magnitudes of the studied co-geoidal signatures, thus effectively limiting the level of detectability of specific structures having specified characteristics; (ii) knowing the spectral characteristics of the altimetry data is the only means that can yield to the design of filtering techniques which make possible the detection of small signals of known form (i.e. salt domes) in the background of the altimetric measurement noise.

The noise spectrum $S_n(f)$ of the altimetry measurement can be conveniently estimated (again) from repeat arcs from the power spectrum density (PSD) of differenced time series, $\zeta_1 \zeta_2$, i.e.

$$S_{\zeta_1 \zeta_2}(f) = 2 S_n(f) \quad (4)$$

where ζ_1 and ζ_2 are the repeat arc segments (Marks and Sailor, 1986).

For the purpose of detection of salt domes or related geological structures, having characterized the spectral model for the altimetry data at a given geographical region, the residual altimetry data series $D(t)$ at time t along a specified subtrack can be modelled as a noise process to which salt dome signatures are added at unknown locations. That is,

$$D(t) = N(t) + SD(t-t_0) \quad (5)$$

where $N(t)$ is a Gaussian random process modelling the altimeter noise, and $SD(t)$ is a salt dome undulation signature centered at sample time t_0 . To detect the presence of a salt dome signature in a sample data set D_S , a statistical hypothesis test is performed; i.e., a null hypothesis H_0 that the salt dome signature is absent, i.e.

$$H_0: D_S(\tau) = N_S(\tau) \quad (6a)$$

is tested against an alternative hypothesis $H(SD, T)$ that a specific salt dome signature occurs at time τ

$$H: D_S(\tau) = N_S(\tau) + SD_S(\tau-t_0). \quad (6b)$$

The correlation analysis of the residual geoid (i.e. after error elimination and geoid regularization corrections) reflected by the satellite altimetry data and the sought salt dome signatures, involves the computation of a linear response function and allows to determine the likelihood that a given signature is present in the sample data set. The theory governing the design and performance of matched filters is well known from the communications and signal processing literature. For this reason we shall refrain from all discussion concerning the derivations and properties of the algorithms and simply outline the steps involved.

First, the Fourier transform $\mathcal{S}D(f)$ of the salt dome signature $SD(t)$ and the power spectral density $S_n(f)$ of the random process that describes the altimetry data noise is used to estimate the transfer function

$$\mathcal{H}(f) = \mathcal{S}D(f) / S_n(f). \quad (7)$$

The second step is to compute the output $\mathcal{D}F(f)$ of the detection filter defined by the Fourier transform

$$\mathcal{D}F(f) = \mathcal{H}(f) \mathcal{D}(f) \quad (8)$$

where $\mathcal{D}(f)$ is the Fourier-transformed track of satellite altimeter data $D(t)$. In the last step, the transformation of the detection filter output in the time-domain is relatively straightforward through an inverse Fourier transform

$$DF(t) = \mathcal{F}^{-1}\{\mathcal{D}\mathcal{F}(f)\} \quad (9)$$

In practice, these steps are naturally implemented following the usual discretization of the Fourier transforms as a consequence of dealing with limited data samples. The performance of the detection filter can be described in terms of the *significant level* α (i.e. the probability α of rejecting H_0 when, in fact, H_0 is true), and the expected *false alarm rate* β (i.e. the probability β of accepting H_0 when, in fact, H_0 is false). Evidently, increasing α (desirable) leads to an increase in β (not desirable), so that a compromise is reached by choosing (usually based on experience) a detection threshold that yields the smallest β -type error for the same significant level α . The likelihood of false alarms is quantified by the statistics of the detection filter itself (e.g. the rms signal to noise ratio assumed by the detection filter output). These may vary from region to region (or even from track to track), the choice of a particular (salt dome) model signature, the data sampling rate and the off-track distance of the given model signature from the closest approach of an altimetry track. Hence, experimental results with known geological structures need to be carefully analyzed before attempting detection of specific model signatures with unknown characteristics in unknown locations. However, this is not considered a serious drawback of the technique considering that the redundancy offered by the numerous altimetry tracks available to date in all parts of the world's oceans can help resolve any ambiguities that one may encounter in the course of such computations.

SUMMARY

In the previous sections, we have outlined in some detail the procedures currently applied by our group to explore the present capabilities of satellite radar altimeter data when used in the area of petroleum exploration. The analysis techniques as described above are being applied to areas of known hydrocarbon producing areas with quite encouraging test results. Present analyses have concentrated in the area between 40° to 50° of the east coast of North America (a region rich in diapiric salt structures and seamounts (giving positive gravity anomalies) both buried within the continental midgeocline and on the abyssal plains beyond the shelf); the Gulf Coast with the submarine Campeche and Sigsbee knolls believed to be due to diapirs; and the North Sea with its wealth of salt diapirs and the fault-bounded graben and igneous centres. Evaluating the available satellite altimetry data in the vicinity of regions of known geologic characteristics currently serves as "ground truth" at this stage of our analysis, and provides additional input to the aforementioned methodology. Complete results of these analyses will be reported in the near future. It is planned that these findings be integrated with the results from other prospecting tools in the endeavour to acquire as much geological understanding as possible of a hydrocarbon prospect.

REFERENCES

- American Geophysical Union (1979) - *Journal of Geophysical Research*, 84 (B8).
- American Geophysical Union (1982) - *SEASAT Special Issue I*, reprinted from the *Journal of Geophysical Research*, 87 (C5).
- American Geophysical Union (1983) - *SEASAT Special Issue II*, reprinted from the *Journal of Geophysical Research*, 88 (C3).
- Berube, M., D. Delikaraoglou and A. Mainville (1988) - "The GEOSAT Altimeter Mean Sea Surface", presented at the Fall Meeting of the American Geophysical Union, San Francisco, Dec.
- Marks, K.M. and R.V. Sallor (1986) - "Comparison of GEOS-3 and SEASAT Altimeter Resolution Capabilities", *Geoph. Res. Lett.* 13, pp. 697-700.
- McConathy, D.R. and C.C. Kirgus (1987) - "The Navy GEOSAT Mission : an overview", *John Hopkins APL, Technical Digest*, Vol. 8, No. 2.
- National Research Council (1979) - "Application of a dedicated gravitational satellite mission", Panel on Gravity and Sea Level, Committee on Geodesy, Washington, National Academy of Sciences.
- Schwank D.C. and A.R. Lazarewicz (1982) - "Estimation of Seamount Compensation Using Satellite Altimetry", *Geophys. Res. Letters*, Vol. 9 (8), pp. 385-388.
- White, J.V., R.V. Sallor, A.R. Lazarewicz and A.R. Schwank (1983) - "Detection of seamount signatures in SEASAT altimetry data using matched filters", *Journal of Geophysical Research*, 88 (C3), pp. 1541-1551.
- Young, T.P., G.F. Lindal and C.H. Liu (1988) - "The Role of GPS in Precise Earth Observation", *Proc. of the IEEE Position Location and Navigation Symposium*, pp. 251-258.

NORCSEX WIND STRESS MEASUREMENTS FROM A SHIP AND A BUOY

K.L.Davidson and C.Skupniewicz
Naval Postgraduate School
Monterey, Ca 93943
Telephone 408 646-2563, Telecopier 408 646-2921

O. Skagseth
Nansen Remote Sensing Center
Bergen, Norway
Telephone 47 5 297288, Telecopier 47 5 200050

Continuous surface layer turbulent wind measurements were made in the NORCSEX region from a forward mast on the R/V Haakon Mosby from 3 to 27 March 1988 and from a buoy for a 40 hour period from 8 to 10 March. Turbulent kinetic energy dissipation rate estimates, derived from variance spectra of hot-film sensed wind fluctuations, were used with data describing the surface layer thermal stability to calculate the wind stress. The friction velocity (directly derived from the wind stress) versus wind speed is the emphasized atmospheric parameter in interpreting recent remote sensing parameters, e.g. Keller et. al (JGR, 1985) and Li et. al (JPO, 1989). Friction velocities were obtained in the NORCSEX remote sensing experiment to examine effects of atmospheric forcing on surface wave modulation mechanisms affecting SAR and Scatterometer backscatter. Thermal stability was one of the factors presented by Keller et. al as being critical to explaining non-linear relationships between scatterometer measured parameters and the surface layer wind speed.

Interpretation of the friction velocity (wind stress) results are on the appropriateness of open ocean derived drag coefficients for the NORCSEX region and on the relationship between ship and buoy measured friction velocities and scatterometer backscatter parameters. Previous coincident "dissipation" derived surface wind friction velocities values and radar (L and ku band) measured parameters were obtained by our group off the US West Coast (Geernaert et. al, JGR, 1988) and in the west North Atlantic (Li et al, JPO, 1989). Those results indicate that friction velocity (wind stress) values are better than wind speed values for explaining radar sensed changes of surface properties. The better correlations with friction velocity (wind stress) than with wind speed for the North Atlantic "open ocean" and west coast "coastal" locations are believed to occur because of sea surface temperature fronts and of coastal influences on the wave field, respectively.

The combined NORCSEX friction velocity and remotely sensed results should yield quite good information, in comparison to the previous results, because of more complete

descriptions of local wave fields and currents and the greater variations in wind forcing intensities. For the NORCSEX region we found the wind speed dependence of the neutral drag coefficient to differ, 20 to 30% larger, from that formulated by Smith (JGR, 1988) and which is often used for open ocean applications. For remote sensing, the relation between friction velocity and backscatter values obtained from a ship mounted scatterometer (X, C and L band) is best described by two linear segments where the subgroups correspond to wind speeds less than 6 m/s and greater than 8 m/s. These results relating backscatter to wind forcing, with a breakpoint near 8 m/s, agree with those presented by Keller et. al.

References:

- Geernaert, G.L., K.L. Davidson, S.E. Larsen and T.M. Mikkelsen, 1988: Wind Stress Measurements During the Tower Ocean Wave and Radar Dependence Experiment, J. Geophys. Res., **93**, 13,913-13924.
- Keller, W.C., W.J. Plant and D.E. Weissman, 1985: The Dependence of X-band Microwave Sea Return on Atmospheric Stability and Sea State, J. Geophys. Res., **90**, 1019-1029.
- Li, F., W. Large, W. Shaw, E. Walsh and K.L. Davidson, 1989: Ocean Radar Backscatter Relationship With Near Surface Winds: A Case Study During Fasinex, J. Phys. Oceanogr., (in press).
- Smith, S.D., 1988: Coefficients for Sea Surface Wind Stress, Heat Flux, and Wind Profiles as a Function of Wind Speed and Temperature, J. Geophys. Res., (in press).

SCATTEROMETER MEASUREMENTS OF WIND, WAVES, AND OCEAN FRONTS DURING NORCSEX

Robert G. Onstott and Robert A. Shuchman
Radar Science Laboratory
Advanced Concepts Division
Environmental Research Institute of Michigan
Ann Arbor, MI 48107 USA

J.A. Johannessen and O. Skagseth
Nansen Remote Sensing Center
Bergen, Norway

K. Davidson
Department of Meteorology
Naval Postgraduate School
Monterey, CA 93940 USA

ABSTRACT

During March 1988, active microwave radar coefficient measurements were made from a ocean going research platform during a wind-wave-current field experiment on the Norwegian Continental Shelf (NORCSEX). Radar backscatter data were collected at 1.5, 5.25, and 9.38 GHz for incidence angles 20° to 80° and with both like and cross polarizations. One of the primary objectives of this study was to investigate the ability of SAR to image ocean surface features caused by current fronts and eddies in the moderate circulation regime off the Norwegian coast. In addition to supporting the SAR imaging aspects of the study, data were acquired for use in the study of the backscatter response of a wind driven sea. Detailed sea truth and meteorological characterizations acquired spatially and temporally coincident with the scattering measurements, and SAR imagery allows the examination of the influence of wind speed, wind direction, and sea state on radar cross section.

1. INTRODUCTION

In preparation for the 1990 launch of the first European Space Agency (ESA) Earth Resource Satellite (ERS-1) which will include a C-band synthetic aperture radar (SAR), a two week wind-wave-current oceanographic field investigation was conducted during March 1988 along the Norwegian Continental Shelf known as the Haltenbanken. A primary objective of this Norwegian Continental Shelf Experiment (NORCSEX) was to investigate the capability of SAR to image ocean surface features associated with current fronts and eddies in the moderate current regime off the Norwegian coast. Secondary objectives of the study were to assess the potential of a C-band SAR to measure ocean surface wind and waves.

In support of the SAR imaging aspects of the study, data were acquired to study the microwave backscatter response of a wind driven sea. A three-frequency four-channel scatterometer operated from the research vessel HAKON MOSBY collected microwave data at frequencies centered at 1.5, 5.25, and 9.38 GHz (L-, C-, and X-band, respectively), with incidence angles from 20° to 80° , and at like and cross linear polarizations. The scatterometer collects microwave data in a real-aperture mode, and thus is not dependent on platform motion (i.e., Doppler effects) as in the case of the SAR.

In this paper the L-, C-, and X-band scatterometer data collected during NORCSEX is first described. Data collected over a meteorological and oceanographic front that was imaged coincidentally with the C-band

SAR is then presented. Exponents relating the scattering coefficients (σ_0) to wind speed are then calculated for the L-, C-, and X-band coincidentally collected radar data.

2. DATA SETS

The location of the March '88 NORCSEX investigation is shown in Figure 1. The remote sensing, meteorological and oceanographic data collected during NORCSEX is summarized in Table 1. The remote sensing data included the ship scatterometer, the C- and X-band SAR, the NOAA satellite Advance Very High Resolution Radiometer (AVHRR) imagery, the GEOSAT satellite altimeter, and DMSP satellite Special Scanning Microwave Imager (SSM/I).

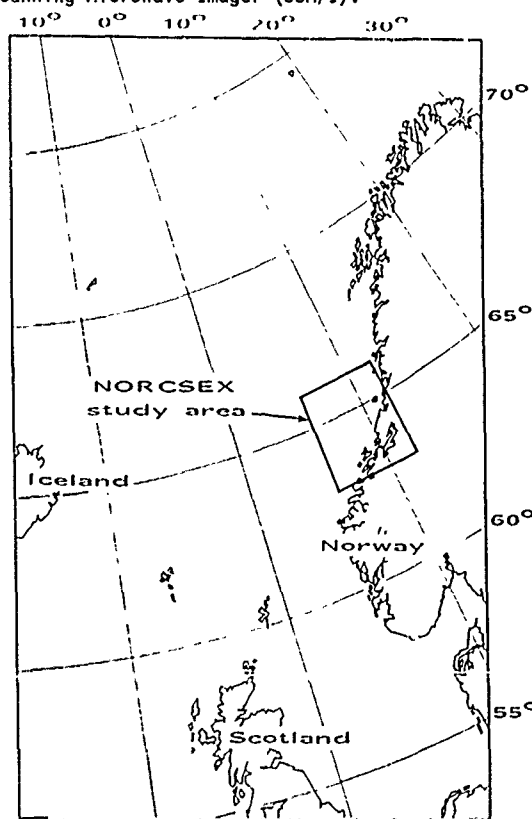


Figure 1. Location of the Norwegian Continental Shelf Experiment Study Area

Table 1. OVERVIEW HALTENBANKEN EXPERIMENT 1988

| PARAMETERS OBTAINED | PLATFORM | TEMPERATURE | | SURFACE | | INTERNAL | FRONTS | NEAR SURFACE | AIR SEA |
|------------------------------|-----------------|---------------|----------|------------|-------|----------|--------|--------------|---------|
| | | OCEAN CURRENT | SALINITY | WIND | WAVES | | | | |
| SHIP MOUNTED WEATHER STATION | ADCP | X | X | | | | X | X | X |
| SEA SOAR | SCATTEROMETER | X (R) | | | | | X | X | |
| COCHR | | X | | | | | | | |
| ELUWYN | WEATHER STATION | X | | | | | | X | X |
| MAR 12 | | | | | | | | | |
| ALCORN | | | | | | | | | |
| CURRENT THERMISTOR | | X | X | X | X | X | | X | |
| WAVES | | | | | | | | X | |
| METEOLOGY | | | | | | | | X | X |
| NOAA 910 | ADCP | | X | | | | X | | |
| GEOSAT | ALTIMETER | X | | X | X | | X (R) | X | |
| CHGP | SSM/I | | | | | | | X (L) | |
| SFT | SLAR | | | | | X | X (R) | | |
| AMCRAFT | SAR | X (R) | | WAVELENGTH | X | X | X (R) | | |

R: Research Evaluation. L: Limitation due to land effect.

The meteorological measurements and data assimilation were conducted during the entire NORCSEX field investigation period. Time series of surface layer meteorological data from ship mounted sensors and profilers of temperature, humidity and vector wind from rawinsondes were obtained. Surface layer wind fluxes (i.e., drag coefficients) were obtained from ship mounted hot-film and the use of miniature cups.

The oceanographic sea truth was obtained by the use of a ship mounted thermistor (sea surface temperature), a towed undulating SeaSoar (salinity and temperature from the surface to 250 m), and the ship mounted 150 KHz Acoustic Doppler Current Profiler (ADCP). The ADCP provides a measure of absolute current every 5 m from the surface to near the ocean bottom.

3. SCATTEROMETER OBSERVATIONS

During this investigation the L-, C-, and X-band radar scatterometer was mounted on the starboard side of the wheelhouse of the R/V HAKON MOSBY (Figure 2). Detailed scatterometer system specifications have been included in Table 2. It was positioned as far forward on the ship as possible and pointed slightly forward of where the ship generated bow waves are produced. Thus, the effects of the bow generated waves were minimized. Bow waves which interacted with the ambient ocean wave field did however, break near the ship creating minimum observation angles of incidence which were sea state dependent. Calm conditions allowed observations at incident angles as small as 20 degrees, where as, in the extreme wave cases minimum angles were between 40 and 50 degrees. Ship roll and pitch motions were recorded as part of the data stream and when they were great enough to cause the backscatter returns to fall outside of the radar intermediate frequency processing filter, these data were noted and not included in the averaging process. To date, 2 and 10 minute averages have been processed. As the ship transited through the oceanographic features and wind fields, four channels of backscatter data and a visual time-encoded video record were acquired.

Radar observations at L-VV, C-HH, X-VV and X-VH or L-HH, C-VV, X-VV and X-VH were made as the ship transited through ocean fronts and internal waves. The first letter refers to the radar frequency used while the VV, HH, and VH denotes the transmit and receive polarizations. For example, VH indicates a cross polarization measurements where vertical polarization was transmitted and horizontal received. The range of meteorological and oceanographic conditions that occurred during the period of scatterometer observations 5 to 18 March are summarized in Table 3.



Figure 2. Three - Frequency, Four - Channel Scatterometer Operated at 1.5, 5.25, and 9.38 GHz, at Angle From 20° to 80°, and Like and Cross Linear Polarization

Table 2. SCATTEROMETER SYSTEM PARAMETERS

| Parameter | L-Band | C-Band | X-Band |
|---|--------|--------------------|--------|
| Frequency (GHz) | 1.50 | 5.25 | 9.38 |
| Wavelength (cm) | 20.0 | 5.7 | 3.2 |
| Polarization | | - VV, VH, HV, HH - | |
| Incidence Angle | | 20° to 80° | |
| Height (m) | | 10 | |
| Spot Size (m) at 40° | 2.1 | 0.7 | 0.7 |
| Bandwidth (MHz) | 350 | 525 | 575 |
| Independent Sampling* Per Meter at 40° | 12 | 6 | 8 |
| Precision* (dB) at 40° | ±1.1 | ±3.0 | ±2.6 |
| Absolute Accuracy | ±1.0 | ±1.0 | ±1.0 |
| Ship Speed (m/s) | | 4 | |
| Look Direction | | Starboard | |

* Assumes the Surface is Stationary

Table 3. Range of Meteorologic and Oceanographic Conditions During Scatterometer Observation Period 5 to 18 March

| | |
|---------------------------|----------------------------------|
| Wind Speed (15 m) | 0 to 25 m/s with gusts to 30 m/s |
| Temperature Air (15 m) | -5° C to +6° C |
| Temperature Sea (Surface) | 2° C to 7° C |
| Waveheight (H 1/3) | 1 to 10 m with 2 - 4 m typical |
| Swell Wavelength | 150 to 300 m |

An example of an ocean front observed on 13 March is shown in the photograph provided in Figure 3. The temperature gradient for this front was +2°C and was created at the boundary between warm Atlantic Ocean water (T=7°C) and the colder Norwegian coastal water (T=5°C). The photograph illustrates the importance of sea surface temperature on the small scale surface roughness. Backscatter angular and polarization response measurements were made on each side of this ocean frontal feature. Quad-polarization measurements (VV, VH, HV, AND HH) were made at L- and C-band and dual-polarization measurements (VV and VH) at X-band. Seven leg star patterns were made to address wave and wind aspect angle dependencies.



Figure 3. Example of an Ocean Front which was Transited on 13 March. The Temperature Gradient was 2 °C. The Front was Created at the Boundary of Warm Atlantic Water (7 °C) and Colder Coastal Water (5 °C).

An interesting data collection took place on 17 March in which fluctuations in wind speed of 2 to 13 m/s were experienced during a ten hour period as the ship transited an oceanographic/meteorological front. Sea temperatures ranged from 5° to 8°C while the air temperature was 3°C. The air-sea temperature difference ($T = T_{air} - T_{sea}$) ranged from -2°C to -5°C producing unstable conditions since $T < -2^\circ \text{C}$. A gravity wave swell traveling in the same direction as the wind with 2-4 m significant wave height was present during the scatterometer observation. The wind and wave direction with respect to the radar was approximately 90° (i.e., upwind/downwind look direction). In Figure 4a, the L-, C-, and X-band radar scattering coefficients measured during this ten hour period are shown as a function of time. The wind speed and direction measured coincidentally with the scatterometer at a height of 15 m is shown in Figure 4b. These figures visually show the correlation between wind speed and radar scattering cross section. Features that are instructive include the general increase in cross section as wind speed increases, the hysteresis associated with wave decay after a reduction in wind speed, and the change in the scattering coefficients of 10-13 dB for an increase in wind speed of 3 m/s to 13 m/s.

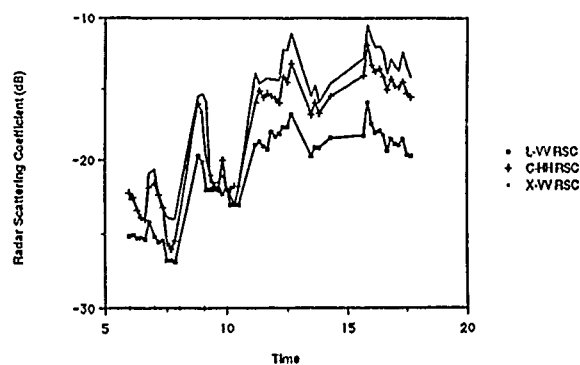


Figure 4a. Radar Scattering Coefficient at L-, C-, and X-band Measured While Crossing an Ocean-Meteorological Front on 17 March

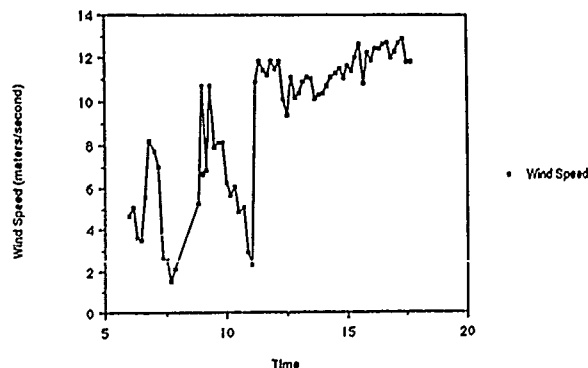


Figure 4b. Wind Speed Measured From Ship While Crossing an Ocean-Meteorological Front on 17 March.

It is also interesting to report that a response identical to that of the C-band scatterometer was observed by the SAR 5.6 GHz data. During this period a rapid transition from a spatially large region of weak returns where winds were calm (<3 m/s) to strong radar returns where the winds greatly increased (>10 m/s) produced a 10 dB difference in backscatter intensity for these two regions [reported by Shuchman et al, 1989, these proceedings]. The ship transited across this front about 30 minutes after the SAR and noted this change in wind speed immediately prior to crossing the ocean front at 11 GMT. The wind friction velocity (U^*) can be related to drag coefficient (C_d) and true wind speed (U) by the following expression

$$U^* = C_d^{1/2} U \quad (1)$$

In this case U^* changed from 0.1 to 0.5 m/s across this front illustrating the important difference in sea surface roughness.

The radar scattering coefficient σ^0 has historically been related to wind speed through a transfer function of the form

$$\sigma^0(U) = C U^\gamma \quad (2)$$

where U is related to the surface wind vector, C the scaling coefficient, and γ the wind vector exponent. In addition to the local wind vector, the radar scattering coefficient can be additionally influenced by a array of other environmental parameters which include wave slope, sea surface temperature, air-sea temperature difference, and surfactants [Donelan and Pierson, 1987; Plant, 1986]. In Figure 5, the radar scattering coefficients at L-, C-, and X-bands are shown as a function of equivalent 10 m (height above the ocean surface) winds. Forty ten-minute intervals were used in producing these plots. All three frequencies produce approximately the same scaling coefficient, but yield different wind vector exponents and correlation coefficients (i.e., they range from 0.6 to 0.8). Exponents increased with increasing frequency and ranged from 1.30 to 1.76.

Keller et al [1989] reported data collected from a tower situated in the North Sea for C-VV and an incident angle of 45° which showed a response very similar to the NORCSEX observations. The radar cross-sections ranged in value from about -24 to -15 dB for a change in wind speed of 2 to 10 m/s. For this same change in wind speed, our scattering coefficients ranged from about -26 to -17 dB. Recall the NORCSEX C-band measurements utilized HH polarization. This change in polarization from vertical to horizontal is sufficient to account for the observed difference. The agreement in absolute level is reasonable and each set of measurements produced a 9 dB change in cross-section for this wind speed difference. The computed wind vector exponent for the results reported by Keller was 1.50. In our case, the exponent is 1.64.

4. SUMMARY

A wide range in meteorological and oceanographic conditions with respect to winds, waves, temperatures occurred during NORCSEX. Several distinct moderate to high winds (greater than 10 m/s) events were encountered as well as low wind conditions (2-4 m/s). Preliminary analysis indicates excellent correlation between variation of observed wind conditions and drag coefficients and variations of shipboard and aircraft remote sensing data. Future efforts will be directed at examining the relationships between wind stress, sea surface temperature, stability, sea and swell on remote wind vector determination.

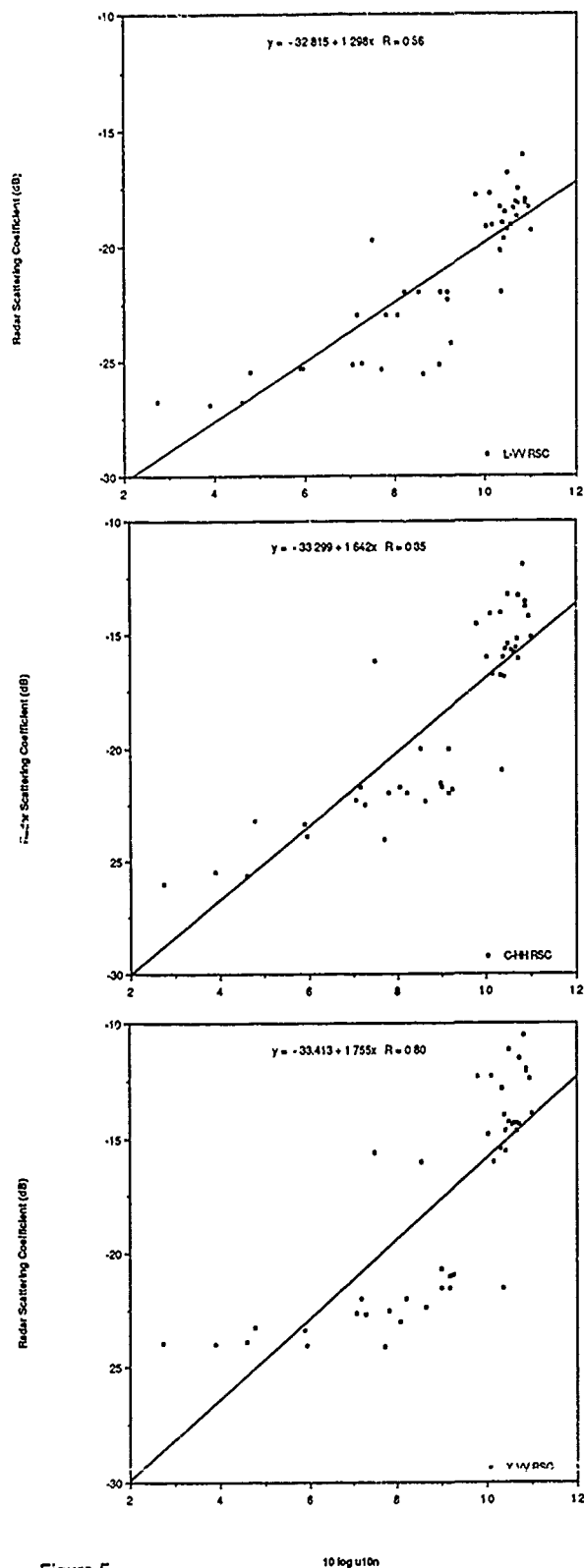


Figure 5

Radar Scattering Coefficients Measured on 17 March at L-VV, C-HH, and X-VV are Shown as a Function of Equivalent 10-m Winds.

ACKNOWLEDGEMENTS

This work was supported by the Oceanographer of the Navy, SPAWAR, and the Office of Naval Research (ONR) under ONR contract #N00014-81-C-0692.

REFERENCES

- [1] Shuchman, R.A., D.L. Lyzenga, J.A. Johannessen, and C. Livingstone, "SAR Detection of Mesoscale Ocean Circulation Features During NORCSEX '88", these proceedings.
- [2] Donelan, M.A. and W.J. Pierson, Jr., "Radar Scattering and Equilibrium Ranges in Wind-Generated Waves With Application to Scatterometry, JGR., 92, pp 4971-5029, 1987.
- [3] Plant, W.J., "A Two-Scale Model of the Short Wind-Generated Waves and Scatterometry, JGR.", 91, pp 10,735-10,749 1986.
- [4] Keller, W.C., V. Wismann, and W. Alpers, "Tower-Based Measurements of the Ocean C-Band Radar Backscattering Cross Section, JGR., 94, pp 924-930, 1989.

CORRELATION OF MARINE RADAR CROSS SECTION STATISTICS
WITH WIND STRESS DURING NORCSEX

Dennis B. Trizna
Radar Division, Naval Research Laboratory
Washington, D.C. 20375-5000

Ken Davidson
Naval Postgraduate School
Monterey, California

ABSTRACT

Calibrated marine radar sea echo samples were collected during NORCSEX on the Haken Mosby research vessel to study the dependence of sea scatter statistics on wind speed and wind stress. Measurements were made aboard ship of wind stress using the dissipation method with hot-wire anemometers. Using recorded ship heading and relative wind direction, the effects of flow distortion due to winds passing across the hull of the ship from the rear relative to the forward mounted instrumentation were eliminated by neglecting data for these geometries. Radar data were collected over a full 360 degrees, within a radius of one kilometer. The peak radar clutter direction was determined from the radar data and a sector 30 degrees either side of this direction was chosen for comparison. This peak clutter direction was not always aligned with the wind, particularly after a change in wind direction, as had been observed previously in FASINEX and in an experiment conducted in the Pacific Ocean. Cumulative distributions of samples were determined for eight different sets of range bins, the center bin of which defined a grazing angle, between 0.35 and 3.5 degrees. Two Weibull model distributions were fitted to each data set, one associated with sea spikes and the second with a distributed scattering mechanism. Median backscatter level for each grazing angle are determined, which is independent of any model fit, as well as sea spike parameters, which have previously been shown to correlate with white cap coverage. Peak clutter directions are compared with wind directions and primary wave directions measured by pitch and roll buoys in the test area.

SCALAR WINDS FROM SSM/I IN THE NORWEGIAN
AND GREENLAND SEAS DURING NORCSEX¹

Per Gloersen
Laboratory for Oceans
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771, USA

Paul Hubanks
Research and Data Systems Corporation
Lanham, Maryland 20706-2106, USA

ABSTRACT

Data acquired with the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) during the Norwegian Coastal Sea Experiment (NORCSEX) in March 1988 have been utilized to estimate scalar winds in the Norwegian and Greenland Seas. The algorithm for calculating these winds was first developed for the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) in order to investigate scalar winds during the Polar Lows Experiment in February 1984; the coefficients in this algorithm have been tuned to accommodate differences in the SSM/I and SMMR instruments.

INTRODUCTION

Passive satellite radiometric data are utilized to calculate scalar surface winds over the polar oceans. The method is applicable at high latitudes because the air is relatively dry, and the corrections for cloudiness and rain are possible without incurring excessive errors. The data source was the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) during the Norwegian Coastal Sea Experiment (NORCSEX) in March 1988. These data were utilized to estimate scalar winds in the Norwegian and Greenland Seas. The algorithm for calculating these winds was first developed (Gloersen et al., 1989) for the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) (Gloersen and Barath, 1978) in order to investigate scalar winds during the Polar Lows Experiment in February 1984.

The SMMR on board the Nimbus 7 was launched in October 1978 for the purpose of determining the utility and precision of obtaining on a global basis information on sea ice coverage, near-surface oceanic winds, sea surface temperatures, atmospheric water vapor, and cloudiness (Gloersen et al., 1984). Also, snow water equivalence and soil moisture were some of the terrain parameters to be studied. All of the algorithms for estimating these parameters made use of the multispectral, dual-polarized radiances measured by SMMR, ten channels in all at the five wavelengths 0.8, 1.4, 1.7, 2.8, and 4.6 cm. In particular, a prelaunch algorithm was developed for obtaining near-surface scalar winds on a global basis (Wilheit et al., 1983). This algorithm makes use of four of the ten SMMR channels, including the 2.8 cm channels with nominal footprint sizes of about 100 Km. While this algorithm works satisfactorily at distances of 300 Km or more from land or sea ice, it is poorly suited for estimating winds in the polar regions near land or the ice edge, and tends to wash out the details of polar low events which are smaller in scale.

The alternative algorithm utilized in this paper was designed specifically for the polar regions which utilizes ratios of a different combination of four channels, at 0.8 and (in the case of SSM/I) 1.55 cm, and subsequently yields valid estimates closer to land and sea ice boundaries. Restricting the estimates to polar regions minimizes the interference from atmospheric water vapor fluctuations. At this point in time, the algorithm for SSM/I has not been tested against a statistically significant data set. This paper represents a status report on the ongoing algorithm development.

PHYSICAL BASIS FOR THE ALGORITHM

Radiation emanating from the ocean surface in the microwave regime can be described by the Rayleigh-Jeans approximation to the Planck radiation law, i.e. the radiated power is linearly dependent on the sensible temperature of the radiating layer. Thus, mi-

¹Support for this work was provided by the Office of Naval Research under reimbursable document No. N0001488WM22006.

crowave radiance is commonly given in units of degrees kelvin and referred to as brightness temperature. The radiance also depends linearly on the emissivity of the radiating surface. For calm seas, the emissivities are typically about 0.5 and 0.2 for the vertically and horizontally polarized channels, respectively (Gloersen and Barath, 1977). When subjected to wind stress, the reflectivity of the oceanic surface decreases from its specular, calm seas value due to the formation of waves and whitecaps. Therefore, the emissivity of a wind-swept ocean increases, resulting in a nonlinear increase in radiance from the ocean with increasing near-surface wind (Hollinger, 1971; Webster et al., 1976). The horizontally polarized radiance increases at a higher rate than the vertically polarized (Gloersen and Barath, 1977).

Estimation of scalar near-surface oceanic winds from observations of microwave radiances is complicated by atmospheric interference arising from fluctuations in atmospheric water vapor and cloud water content. It is further complicated by variations in the ocean surface temperature in two ways, by the direct dependence mentioned earlier and by a temperature dependence on the onset of whitecapping (Monahan and O'Muircheartaigh, 1986). All of these complications are minimized by restricting these estimations to polar waters, where the ranges of sea surface temperature and the atmospheric water vapor are small. Variation in cloud water content is also generally smaller in the polar regions, but must be taken into account when estimating winds.

FUNCTIONAL FORM OF THE ALGORITHM

In order to minimize further the dependence of wind estimation from microwave radiance on oceanic surface temperature, the microwave polarization, defined as the ratio of the difference of the vertically and horizontally polarized radiances at a given wavelength and their sum, is used as the independent variable. In this way, the effect of sea surface temperature is eliminated to first order. As has been shown previously (Gloersen et al., 1989), this approach has the additional advantage of a linear relationship between the oceanic scalar winds and the observed polarization. Either the polarization at the 0.8 or 1.55 cm wavelength could be used for these estimates, but the longer of the two is less subject to interference from clouds and is therefore selected. Since the sensitivity of polarization to wind is about the same at either wavelength, but not the same for cloud water, the difference in the polarizations at the two wavelengths is used to detect cloud water amount.

Defining the polarization at 1.55 cm as PR and the difference in the polarizations at 0.8 and 1.55 cm as DP, estimates of the near-surface scalar winds, W, and cloud liquid water amount, L, are obtained as follows:

$$W = W_1*(PR - W_0) + W_2*(DP - L_0) \text{ (knots)} \quad (1)$$

$$L = L_1*(PR - W_0) + L_2*(DP - L_0) \text{ (cm)} \quad (2)$$

where, tentatively,

$$\begin{array}{ll} W_0 = 0.242 & L_0 = 0.056 \\ W_1 = -806.4 & L_1 = -0.217 \\ W_2 = -618.3 & L_2 = 0.499 \end{array}$$

The above coefficients are the same as those for SMMR (Gloersen et al., 1989). The justification for tentatively using this set of coefficients is that when overlapping data from SMMR and SSM/I during July-August 1987 were compared on a DP, PR scatter diagram, the points were found to fall in essentially the same area and with similar patterns. Ultimately, the same statistical procedure used to obtain the SMMR coefficients by comparison of SSM/I data with in situ data (Gloersen et al., 1989) will be used to obtain refined coefficients for SSM/I.

ILLUSTRATIONS OF SURFACE WIND AND CLOUD WATER ESTIMATES FROM SMMR

In lieu of SSM/I images from NORCSEX which are not available as of this writing, we show in Figures 1-2 grid-print maps of near-surface oceanic scalar winds and cloud liquid water content in the Norwegian, Greenland, and Barents Seas for the ascending nodal pass of the SMMR on 27 February 1984, which occurs at approximately 1000 GMT. The polar low observed at 69°N and 3°W by Shapiro et al. (1987) can be seen as well in Figures 1 and 2 at the same location. In both figures, the '*' points to the center of the storm, located near the edge of the SMMR orbital swath. (The wedge-shaped data gap is the space between adjacent swaths.) Winds ranging from 30-70 knots can be seen in this vicinity. An area of strong wind extends all the way from the sea ice edge near Greenland to the coast of Norway. This is a marked change from the situation 9 hours earlier when the winds were generally weaker. The cloud patterns (Figure 2) are in the form of circular bands to the east of the storm center, with one band just off the coast of Norway and another about 300 Km to the west. They are approximately centered on the polar low.

SUMMARY

Multispectral microwave radiances obtained from the SSM/I have been compared to SMMR data during a period when both instruments were in operation. As a result, it was found that the algorithm coefficients for the SMMR would serve as a satisfactory initial set of coefficients for SSM/I. An example of a polar low observed with the SMMR is shown to illustrate the technique. SSM/I data showing two high wind events during NORCSEX will be described at the IGARSS'89 symposium.

ACKNOWLEDGMENT

The authors wish to thank Dr. Paul F. Twitchell for useful discussion and encouragement.

REFERENCES

1. Gloersen, P., E. Mollo-Christensen, and P. Hubanks, "Observations of Arctic Polar Lows with the Nimbus 7 Scanning Multichannel Microwave Radiometer", accepted for Polar/Arctic Lows '88, Paul Twitchell, Ed. (A. Deepak Publ.), 1989
2. Gloersen, P., D. J. Cavalieri, A. T. C. Chang, T. T. Wilheit, W. J. Campbell, O. M. Johannessen, K. B. Katsoros, K. F. Kunzi, D. B. Ross, D. Staelin, E. P. L. Windsor, F. T. Barath, P. Gudmandsen, E. Lanham, and R. O. Ramseier, "A Summary of Results from the First Nimbus 7 SMMR Observations", *J. Geophys. Res.*, **89**, 5335-5344, 1984.
3. Gloersen, P., and F. T. Barath, "A Scanning Multichannel Microwave Radiometer for Nimbus-G and SEASAT-A", *IEEE J. Oceanic Eng.*, **OE-2**, 172-178, 1977.
4. Monohan, E. C. and I. G. O'Muircheartaigh, "Whitecaps and the Passive Remote Sensing of the Ocean Surface", *Int. J. Remote Sensing*, **7**, 627-642, 1986.
5. Shapiro, M. A., L. S. Fedor, and T. Hampel, "Research Aircraft Measurements of a Polar Low over the Norwegian Sea", *Tellus*, **39A**, 272-306, 1987.
6. Webster, W. J., T. T. Wilheit, D. B. Ross, and P. Gloersen, "Spectral Characteristics of the Microwave Emission from a Wind-Driven, Foam-Covered Sea", *J. Geophys. Res.*, **81**, 3095-3099, 1976.
7. Wilheit, T. T., J. Greaves, D. Han, B. M. Krupp, and A. S. Milman, "Retrieval of Ocean Surface Parameters from the Scanning Multichannel Microwave Radiometer on the Nimbus-7 Satellite", *IEEE Trans. GRS*, **GE-22**, 133-143, 1984.

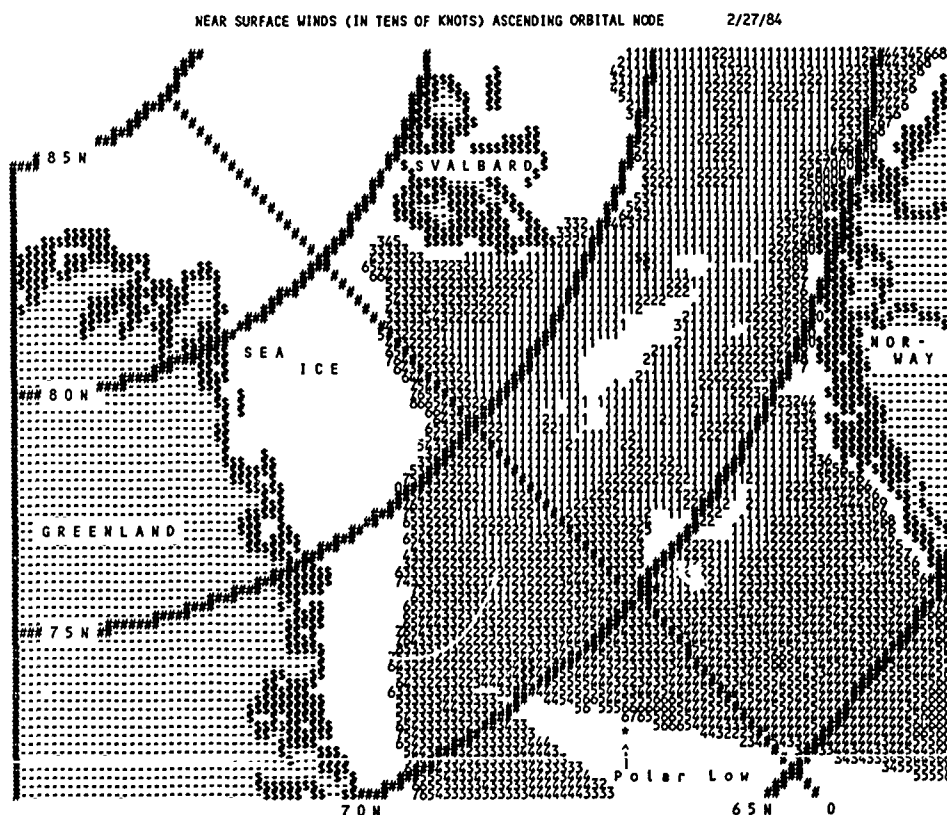


Fig. 1. Near-surface wind estimates from the Nimbus-7 SMMR. The winds are indicated in 10-knot intervals, starting with '1' = 0-10 knots. '0' indicates 90-100 knots. The time of passage over 0°E, 70°N was about 1000 GMT

CLOUD LIQUID WATER CONTENT (IN UNITS OF 0.01 CM) ASCENDING ORBITAL NODE

2/27/84

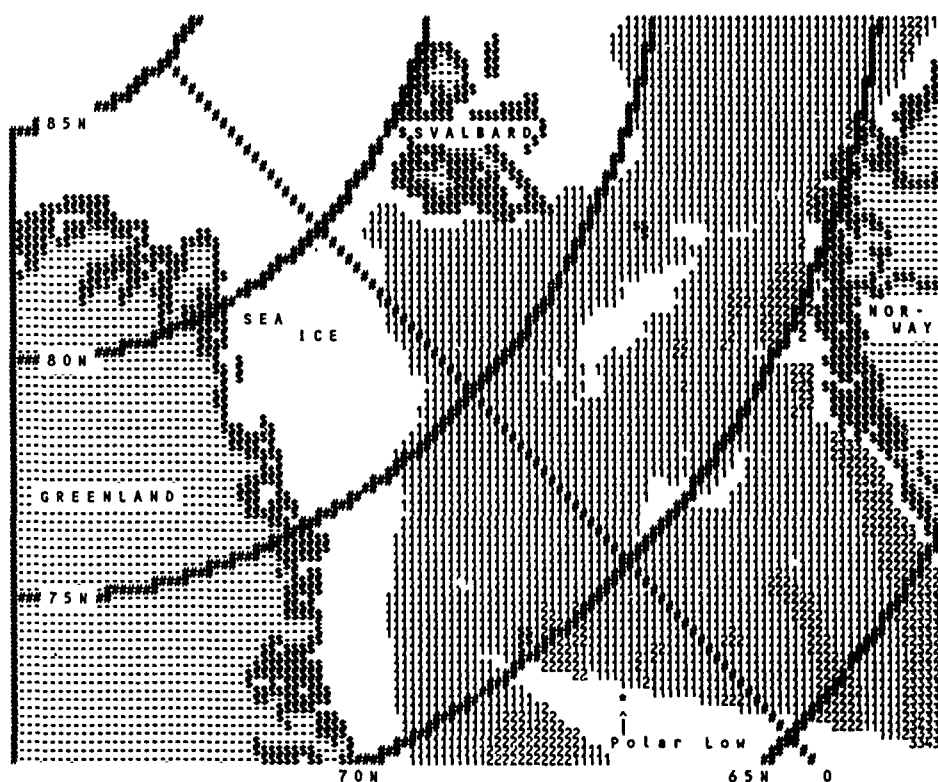


Fig. 2. Estimates of liquid water content of clouds from the Nimbus-7 SMMR. The water is indicated in 0.01-cm intervals, starting with '1' = 0-0.01 cm. '0' indicates 0.09-0.10 cm. The time of passage over 0°E, 70°N was about 1000 GMT.

EVALUATION OF GEOSAT ALTIMETER WIND AND WAVE RESULTS

Lars-Anders Breivik

The Norwegian Meteorological Institute
P.B.320 Blindern 0314 Oslo 3 Norway

ABSTRACT

A system for evaluation of sea surface satellite data has been constructed. GEOSAT altimeter wave height and wind measurements from the North Sea, Norwegian Sea and Barents Sea, March 1988, has been compared to NORCSEX'88 buoy measurements and calculations from numerical models. The tested GEOSAT wind data were found to be of variable quality. As expected the wave height data proved to be of better quality and might be used as initial data in numerical models.

KEYWORDS: altimeter, wind speed, wave height, evaluation

1. INTRODUCTION

A major limitation for improvements of current atmospheric and oceanographic numerical forecasts is the lack of observed data over the oceans. This may be changed by the use of satellite sea surface data.

Compared to conventional meteorological in situ measurements the satellite wind and wave information differs in representativity both in time and space. Such information has therefore to be validated before operational use can be prepared. In this work we are studying altimeter data from GEOSAT, and compare it with in situ measurements from the NOCSEX-88 field experiment at Haltenbanken and with DNMI's model analysis. The investigation period is March 1988. The objectives are:

- 1: To build up an evaluation system handling satellite data.
- 2: To evaluate the three types of data especially considering potential use of altimeter data in model assimilation.

2. DESCRIPTION OF THE MEASUREMENT SYSTEMS

The data is collected from three independent sources of information: observations from buoys and satellite and numerical model analysis. In this chapter we will describe the three different data systems.

2.1 Model analysis

DNMI are running two limited area atmospheric models operationally (Grønås, Hellevik 1982). LAM150 is covering the Atlantic side of the northern hemisphere with a grid point distance of 150 km. LAM50 is a mesoscale model with a grid point distance of 50 km (fig.2.1). Both in LAM150 and LAM50 11 pressure levels are used for analysis and 10 sigma levels are used in the prediction model. The models are developed by scientists at DNMI.

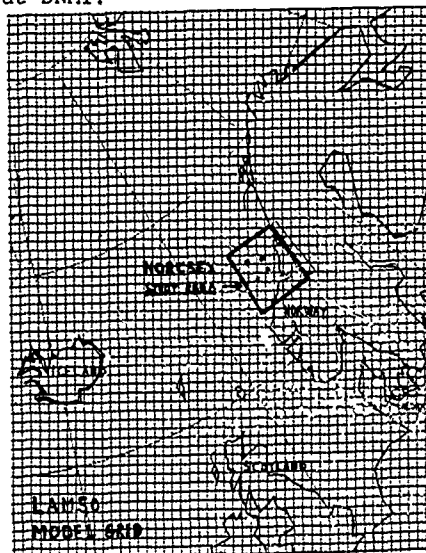


Fig. 2.1 LAM50 integration area.

To create the initial field the models use a method of multivariate analysis. (Grønås, Midtbø 1986). The basis of the analysis is a great amount of meteorological data from an international network of observations. The idea of the analysis is to assimilate the observations to the model grid points. In the grid points the observations are weighted against a first guess field from the previous 6 hours model prognose.

The model winds used in this study are analysed winds from the LAM50 model reduced from the lowest model level to 10 m above sea surface.

A numerical spectral wave model is operational at DNMI. It is called WHINCH and is developed by Oceanweather Inc., Cob, USA. A description of the model and an evaluation of its performance is contained in the SWAMP report (SWAMP 1985) where it is referred to as the SAIL model. The model produces a full two dimensional wave spectrum at 15 frequency bands (0.04 - 0.24 Hz) and 24 directional sectors (15 width). The time step is two hours and the grid resolution is 150 km in the Atlantic and 75 km in the North Sea, Norwegian Sea and the Barents Sea. The operational model input is friction wind velocity from LAM50 predictions. No assimilation of wave observations are involved so far. However, for the period of this investigation, the model was rerun and updated using LAM50 wind analysis. These results will be referred to as model analysis in this report.

2.2 Buoy measurements

During NORCSEX insitu measurements from 4 moored metocean data buoys, deployed by Oceanor, were collected (ODAP 1988). It was one WAVESCAN (Station 1) and 3 NORWAVE data buoys (Station 2, 3, 4). The location of the NORCSEX study area and the positions of the 4 buoys are shown in fig. 2.1 and 2.2. A specification of the measurement systems is given in table 2.1. The buoys measure the wind using an instrument 4 meter above the sea surface. The NORCSEX buoy measurements are not among the observations used in the model analysis.

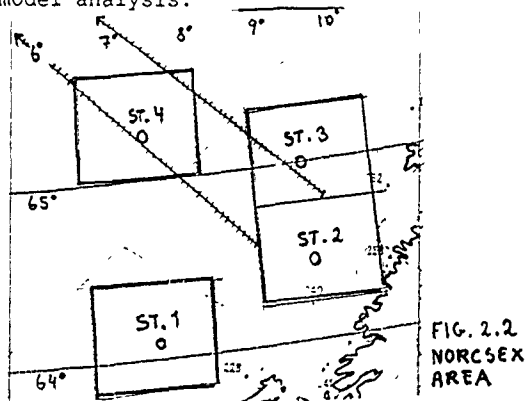


FIG. 2.2
NORCSEX
AREA

| | WIND SPEED | WAVE HEIGHT |
|---------------------|--|---|
| TYPE OF SENSOR | 3 Cup Anemometer | Wave Sensor accelerometer |
| MEASURING TECHNIQUE | Averaged over a 10 minute integration period | Double integrated accelerometer signal |
| SENSOR LOCATION | WAVESCAN: 4.2 m NORWAVE: 3.2 m - above sea level | Near the buoy's centre of buoyancy |
| RANGE | 0.5 to 46.8 m/s | WAVESCAN: +/- 10 m NORWAVE: +/- 15 m |
| ACCURACY | WAVESCAN: +/- 4 % NORWAVE: +/- 5 - 10 % | WAVESCAN: < 3.0 % NORWAVE: < 5 % |

Tab. 2.1 Sensor specifications Ref. 3

2.3 Satellite measurements

We are studying wind and wave data from GEOSAT. GEOSAT is operated by the US-Navy and has been successfully on duty since March 1985. It is equipped with a radar altimeter. A radar altimeter is an active microwave sensor collecting information about reflection time and scattering of transmitted micro waves. The collected raw data are analysed and processed by NOAA. The satellite and the processes involved is described in NOAA Technical Memorandum NOS NGS-46 (NOAA 1987).

The altimeter transmits electromagnetic pulses and measures the backscattered returns from the sea surface. The wind is retrieved by the smoothed Brown algorithm based on the principle that for increasing winds the sea surface becomes increasingly roughened and less radiation are scattered back to the altimeter. The altimeter wave heights are based on the measured time for the return signal of the radar.

DNMI received, on tape from NOAA, USA, processed wave data covering all oceans for the period of March 1988. These data are used for a comparison with the wave model calculations from the North Sea, Norwegian Sea and Barents Sea. From Nansen Remote Sensing Center, Bergen, we received processed wind and wave data covering the NORCSEX experiment area which we used for comparison with both model calculations and buoy measurements.

3. THE DATA

Before the comparison the data from the three different sources has been checked to remove obvious errors from the data sets.

3.1 The model analysis

The observations used in the LAM50 atmospheric model analysis are subject to internal checks prior to the analysis. An additional quality control is performed as a part of the analysis scheme. This is a gross control to get rid of observations with obvious errors. Each observed parameter is checked against the background field. The deviations from the background field are not allowed to extend certain rather wide limits.

3.2 The observations

The processed results from the buoys is of variable quality and contain some obvious errors. First part of the quality control is to introduce maximum and minimum values. As a background field we used the model analysis. Reasonable deviations are allowed. Especially considering the wind observations we might expect variations which is not captured in the model.

The second part of the control is an evaluation of the gradients in the measured time series. For significant wave height we do not allow greater changes than 1 m pr. hour. The basic effect of the quality control is to remove great spikes from the series.

The quality control of the GEOSAT altimeter data, performed before the distribution, is described in NOAA Technical Memorandum.

4. PRESENTATION AND DISCUSSION OF RESULTS

4.1 Wind

First we will compare the model analysis with buoy measurements. Fig.4.1 shows a scatter diagram for station 2. (map: fig. 2.1). Tab.4.1 is a table showing some statistical parameters comparing the two data sets. The table contain the following parameters:

NUMBER: Number of cases compared.

MEAN 1,2: Mean value of the data, number 1,2 referring to the head line.

BIAS: Mean value of the difference: dataset 1 - dataset 2.

RMS: Root mean square difference.

SD: Standard deviation of the differences.

The best results come from station 2 where the root mean square difference is 2.6 m/s, but the analysed winds are systematically lower than the buoy winds. The results for station 1 are also reasonable with an rms-difference equal 3.5 m/s. But here the situation is changed and the analysed wind speeds is heigter than the buoy measurements.

At station 3 and 4 the results are poorer, station 3 having a positive bias and station 4 having a negative one.

From a model point of view the four stations are situated close to each other (ca. 60 -120 km). If the differences between the model and the buoy results were caused by a systematic bias in the analysis, we could, on the average, expect to find very simular difference distributions. But since this is not the case, we assume that an important part of the differences is due to the four independent instruments at the buoys. The accuracy of the wind measuring instruments are for the Wavescan buoys stated to be +/-4% (tab.2.1). This is probably very optimistic. An interesting experiment would be to deploy the buoys at the same location at the same time and compare the wind measurements.

Now we will study the GEOSAT altimeter wind measurements. The GEOSAT near surface wind speed is processed by the use of the smoothed Brown algorithm. We will look at the situations when the satellite passed over Haltenbanken close enough to be compared with measurements from one of the four buoys. Fig.2.2 shows the area and two satellite ground tracks. When part of the track passes through one of the squares surrounding the stations we calculate the mean value of the measurements inside that square, and compare it with the buoy measurement and a model analysis. The results of the comparison are shown in tab.4.2 and on the scatter diagrams in fig.4.2. The number of cases are relatively small, but still, the results seems to be quite bad. The root mean square difference is 5.1 m/s compared with the model analysis and 5.9 m/s compared with buoy measurements. The bias is positive in both cases showing that the altimeter underestimate the wind, but a standard deviation of the errors of about 5 to 6 m/s indicate small chances to find systematic errors. Here is clearly room for improvements.

4.2 Wave

We will start with comparing the model analysis and the buoy measurements.

Time series from station 1 and 2 are shown in fig.4.3. It seems like the model tend to overestimate the small and normal wave heights and underestimate the large ones. This is confirmed in table 4.3. The meanvalue of the model wave height is 0.4 m above the observed.

To compare the three systems together we pick out observations in the same way as for the wind. Results are shown in tab.4.4 comparing buoy and altimeter observations showing good agreement. The bias is close to 0, and the root mean square error is ca. 1 m.

Comparing satellite results with stationary observations has the limitation that it gives few number of cases. This problem is removed comparing the satellite results with results from a model. We have compared the model analysed wave heights with the altimeter results from the North Sea, Norwegian Sea and the Barents Sea for the hole investigation period.

Looking at the total number of 3614 cases we can see the results of the comparison on the scatter diagram and table of fig.4.4 and tab.4.5. The results confirm what we found comparing the data from Haltenbanken. According to the altimeter the wave model overestimate the significant wave height. The bias is +0.6 and rms is 0.9.

5. CONCLUSIONS

During this work we have built up a system for evaluation of data from different kind of information systems. Field data in grid and of regular time steps are compared with in situ measurements averaged in time and with satellite data averaged in space, both types in irregular time intervals.

The main object of the data evaluation is to prepare for assimilation of satellite sea surface data in numerical forecast models. The altimeter wind data tested in this study are to bad to be used for this purpose. The evaluation against buoy measurements show that the model analysed winds are already far more reliable.

The altimeter wave results we have tested shows, not surprisingly, to be of good quality. Both the buoy measured and the altimeter wave heights indicate that the wave model overestimate the actual wave heights. This leads to the conclusion that altimeter measurements might be assimilated in the wave model with a positive influence. Such experiments have been carried out several places (Jansen 1988).

The analysed winds are interpolated from lowest pressure level in the model to 10 meter above sea level. The altimeter is measuring the sea surface roughness. From an assimilation point of view this can turn out to be a more interesting parameter. The roughness is related to the friction wind and from this the actual wind is derived. To do this one need knowledge about the boundary layer. But the friction wind can also be directly used as input to the wave model. This aspect is even more interesting considering the ERS-1 scatterometer sea surface roughness and winds.

REFERENCES

- 1: Grønås, S. and Hellevik, O.E., 1982. A limited area prediction model at the Norwegian Meteorological Institute. Techn. Rep. No. 61, The Norwegian Meteorological Institute, Oslo.
- 2: Grønås, S. and Midtø, K.H. 1986. Four dimensional data assimilation at the Norwegian Meteorological Institute. Techn. Rep. No. 66, The Norwegian Meteorological Institute, Oslo.
- 3: SWAMP (Sea Wave Modeling Project) Group (24 authors), Ocean Wave Modeling, Plenum Press, New York (1985).
- 4: Barstow, S.P. and Bjerkén, S., 1988. Wave measurements from moored directional wave buoys during NORCSEX'88. ODAP Report no. 108, OCEANOR, Trondheim.
- 5: NOAA 1987. GEOSAT Altimeter Geophysical Data Record User Handbook. NOAA Technical Memorandum NOS NGS-46.
- 6: Jansen, P., Lionello, P., Reistad, M., Hollingsworth, A., 1988. A study of the feasibility of using sea and wind information from ERS-1 satellite. ECMWF report to ESA.

STATION 2. 64.1 N. 9.5 E
WIND SPEED M/S
NUMBER OF CASES: 288

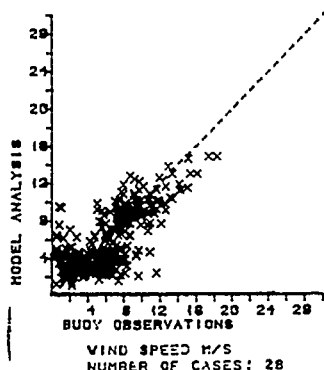


FIG. 4.1

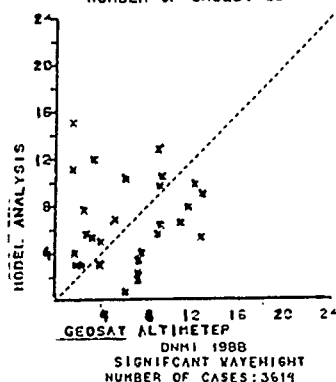


FIG. 4.2

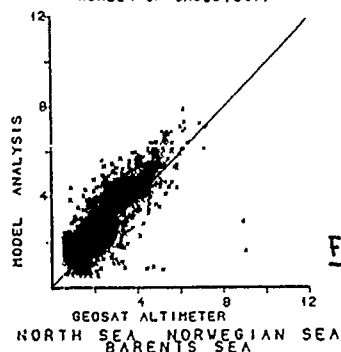


FIG. 4.4

| | | | | | | | | |
|-------------|---------|----------------------|---------|-------|------|-----|----------|--|
| COMPARISON: | | 1: MODEL ANALYSIS | | | | | TAB. 4.1 | |
| | | 2: BUOY OBSERVATIONS | | | | | | |
| PARAMETER: | | WIND SPEED M/S | | | | | | |
| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: | | |
| STATION 4 | 288 | 7.0 | 8.3 | -2.6 | 3.3 | 3.5 | | |
| STATION 3 | 208 | 6.8 | 5.9 | 0.7 | 3.1 | 4.4 | | |
| STATION 2 | 288 | 5.9 | 6.3 | -0.3 | 2.6 | 2.5 | | |
| STATION 1 | 288 | 5.9 | 5.9 | 1.1 | 3.5 | 3.3 | | |

COMPARISON:

1: MODEL ANALYSIS

2: GEOSAT ALTIMETER

TAB. 4.2

PARAMETER: WIND SPEED M/S

| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: |
|------------|---------|---------|---------|-------|------|-----|
| NORCSEX 88 | 28 | 6.9 | 6.7 | 0.3 | 5.1 | 5.1 |

COMPARISON:

1: BUOY OBSERVATIONS

2: GEOSAT ALTIMETER

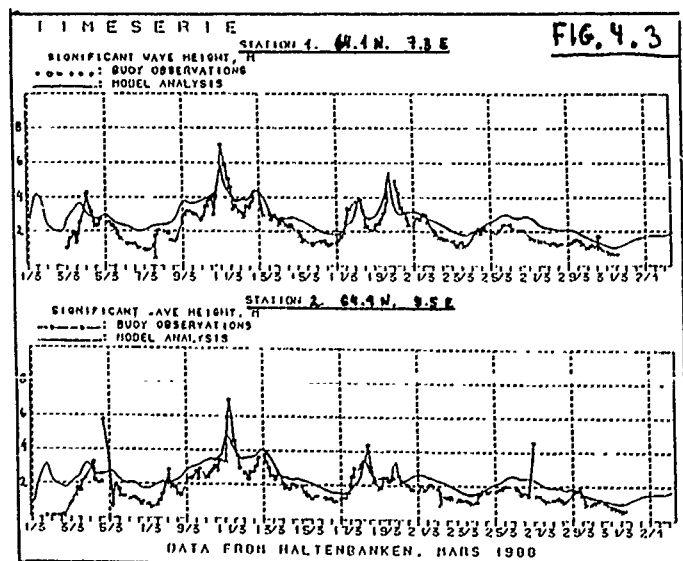
PARAMETER: WIND SPEED M/S

| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: |
|------------|---------|---------|---------|-------|------|-----|
| NORCSEX 88 | 21 | 8.3 | 6.7 | 1.5 | 5.9 | 5.7 |

| | | | | | | | |
|-------------|---------|----------------------------|---------|----------------------|------|----------|--|
| COMPARISON: | | 1: MODEL ANALYSIS | | 2: BUOY OBSERVATIONS | | TAB. 4.3 | |
| PARAMETER: | | SIGNIFICANT WAVE HEIGHT, M | | | | | |
| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: | |
| NORCSEX 88 | 1114 | 2.6 | 2.2 | 0.7 | 0.9 | 0.5 | |

| | | | | | | | | |
|---------------------------------------|---------|----------------------|---------|-------|------|-----|----------|--|
| COMPARISON: | | 1: BUOY OBSERVATIONS | | | | | TAB. 4.4 | |
| | | 2: GEOSAT ALTIMETER | | | | | | |
| PARAMETER: SIGNIFICANT WAVE HEIGHT, M | | | | | | | | |
| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: | | |
| NORCSEX 88 | 21 | 2.2 | 2.3 | -0.0 | 1.0 | 1.0 | | |

| | | | | | | | |
|-------------|---------|----------------------------|---------|---------------------|------|----------|--|
| COMPARISON: | | 1: MODEL ANALYSIS | | 2: GEOSAT ALTIMETER | | TAB. 4.5 | |
| PARAMETER: | | SIGNIFICANT WAVE HEIGHT, M | | | | | |
| | NUMBER: | MEAN 1: | MEAN 2: | BIAS: | RMS: | SD: | |
| NORCSEX 88 | 3614 | 2.8 | 2.3 | 0.6 | 0.9 | 0.8 | |



C-BAND SYNTHETIC APERTURE RADAR MEASUREMENTS OF MOVING OCEAN WAVES

Clifford L. Rufenach
 Albert O. Scheffler
 Robert A. Shuchman
 Radar Science Laboratory
 Environmental Research Institute of Michigan
 Ann Arbor, MI 48107 USA

Johnny A. Johannessen
 Nansen Remote Sensing Center
 Bergen, Norway

ABSTRACT

SAR observations taken during the NORCSEX '88 experiment were specially processed to investigate the distortion of ocean wave imagery caused by surface motion. The special processing included a one-dimensional spectral analysis to obtain the azimuth cutoff for a variety of R/V ratios and surface winds. Large changes in surface winds were available (2 to 10 m/s) across an atmospheric front. The rms facet velocity σ_v was calculated for waves traveling in the range direction for which the velocity bunching is small; $\sigma_v \approx 0.2$ to 0.3 m/s for C-band resolution of 10 and 20 m in azimuth and ground range. Indeed, these measurements show that the rms facet spatial width σ_x varies linearly with R/V and is independent of surface wind velocity for the limited data analyzed. The linear dependence with R/V is expected while the independence of wind velocity is unexpected. Furthermore, it is suggested that the relevant spatial size over which the facet velocities contribute to smearing is the degraded radar resolution rather than the stationary resolution.

I. INTRODUCTION

Synthetic aperture radar (SAR) measurements from both spaceborne and airborne platforms have been used to estimate directional wave spectra [Ref. 1-3] including the peak wavenumber and direction. These SAR measurements are in agreement with surface wave measurements except at high sea states and/or short wavelengths when the wave spectrum is sometimes distorted with the ocean wavenumber vector rotated toward the radar range direction. The radar backscattering is given by the two-scale Bragg scattering model. This model means that the backscattered field cannot be represented by individual infinitesimal scattering sources but rather by the resonant return at a single ocean wavenumber. The smallest area over which this resonance can occur is called a facet which is small compared to the long ocean wavelength. Long ocean waves are imaged by SAR through a combination of mechanisms, including surface slope, surface roughness, and surface motion. Each of these mechanisms can be described by a linear modulation transfer function over a limited range of ocean/radar conditions. This linearity allows an estimate of the directional wave spectrum from the image spectra in a straight-forward manner. However, it is now generally agreed that non-linearities due to the surface motion [Ref. 4] often complicate the extraction of the waveheight spectrum.

This motion can be characterized as degrading the azimuth resolution due to two effects: (1) non-linear mapping of the resolution cells' positions, and (2) sub-resolution smearing. The non-linear mapping is caused by orbital motion of waves longer than the resolution size and is usually called non-linear velocity bunching. The smearing is caused by differential random facet velocities for sizes smaller than the resolution cell. Orbital acceleration of the long waves has been shown to be a second order effect [Ref. 5] under most conditions, and it is not considered here.

The airborne C-band SAR observations off the Norwegian coast (NORCSEX '88) during March, 1988 are used to investigate these motion effects. The NORCSEX SAR measurements were taken over a variety of wind and wave conditions with surface measurements available in the same area. The primary motivation for this experiment was to better understand the capability of C-band SAR in detecting ocean surface features. This understanding can then be applied to the spaceborne C-band SAR scheduled for launch on the European Space Agency spacecraft, ERS-1. The NORCSEX imaging radar measurements were taken at both C-band and X-band with multi-sided flight patterns in either narrow or wide swath mode. Furthermore, the measurements were processed for a variety of range to platform velocities (R/V ratios) across the radar swath width including the near range. The near range R/V = 110 is the same R/V as ERS-1.

Beal et. al. [Ref. 1] reported a cutoff linearly dependent on the R/V ratio and a square root dependence on $H_{1/3}$; whereas Alpers and Bruening [Ref. 5] and Tucker [Ref. 6] reported a surface wind speed dependence under certain conditions in addition to the R/V and $H_{1/3}$ dependencies. These variations were based on an analytical model which is a function of the spread in radial facet velocities within a radar resolution cell. They did not include a non-linear velocity bunching dependence since no analytical expression is available.

The degraded radar resolution can be studied by special processing of the images into azimuth spectra. This processing includes spectral analysis, intensity normalization, range averaging, and a system impulse correction. The resulting spectra are displayed as a function of azimuth wavenumber from which the spectral width or equivalently the high wavenumber cutoff is obtained. This width is dependent on both the velocity bunching and velocity smearing. The variation of this spectral width was investigated as a function of two parameters: R/V ratio and the surface wind speed.

II. EXPERIMENT DESCRIPTION

NORCSEX '88 was a carefully planned experiment containing a number of remote sensing and in-situ instruments. The remote sensing instruments included the Canadian CV-580 aircraft configured with a dual wave-length SAR [Ref. 7], altimetric overflights using the GEOSAT spacecraft, and a ship mounted scatterometer. The surface wind and wave measurements were acquired from four pitch and roll buoys and two research ships. The aircraft SAR acquired data for a total of 28 hours during 6 dedicated days between March 11 and March 28, 1988 with vertical transmit and receive polarizations and a seven-look real-time digital processor. During NORCSEX, gravity waves varied in waveheight from 1 to 10 m with typical heights from 2 to 4m and wavelength of 150 to 300 m. The wind varied from 0 to 25 m/s. The radar operated in two modes: (1) narrow swath with a range width of 16.4 km, and (2) wide swath with a width of 63 km. The measurements of interest in the present work were primarily observed in the wide swath mode. The multi-look radar resolution is 5.6 m by 6 m in slant range and azimuth in the narrow mode and 20 m by 10 m ground range and azimuth in the wide mode. The radar signals are digitally recorded with 4096 range cells for each mode.

III. DATA ANALYSIS

The SAR digital tapes were furnished by the Canadian Centre for Remote Sensing. The image intensity spectra and the azimuth scans of these spectra were processed at the Environmental Research Institute of Michigan. The one-dimensional azimuth spectra were obtained by averaging 64 range pixels selected near the peak wavenumber. The SAR image spectra were calculated from 512 x 512 samples (pixels). The spatial sampling is 15.7 m in ground range and 6.22 m in azimuth and the analysis area is 4.019 km by 1.592 km. Therefore the Nyquist wavenumbers are .200 rad/m and 0.505 rad/m in ground range and azimuth. All parameters are for the wide mode unless otherwise indicated.

The azimuth spectra were obtained by: first normalizing the image intensity by calculating its zero mean divided by its mean, $(I - \langle I \rangle) / \langle I \rangle$, secondly, performing the range averaging, and thirdly, correcting for the impulse response of the radar. The resulting azimuth spectra were plotted on both semi-log and log-log coordinates. Figure 1 illustrates the semi-log variation for the near and far edge of the swath. The spectra are described by a low wavenumber constant response out to a break point ($k_x \approx 0.015$ rad/m and $\sigma_k \approx 0.025$ rad/m for the far and near range, respectively) followed by a spectral decrease out to another break point ($k_x \approx 0.03$ - 0.06 rad/m) followed by a high wavenumber noise floor. This is a typical response which can be represented by a Gaussian low-pass filter $F(k_x)$ for $k_x < k_n$.

$$|F(k_x)|^2 = \exp(-k_x^2 / \sigma_k^2), \quad (1)$$

where $\sigma_k = 1/\sigma_x = (V/R)/\sigma_v$ is the rms spectral width, σ_x is the rms spatial width. The width σ_v is the rms radial facet velocity within a degraded resolution cell given by,

$$\sigma_v^2 = \int_{\Omega_{SAR}} G^2(\theta, \phi) S_w(\Omega) \Omega^2 d\Omega, \quad (2)$$

where $S_w(\Omega)$ is the one-dimensional waveheight spectrum, $\Omega_{SAR} = (g k_{SAR})^{1/2}$, k_{SAR} is the inverse azimuth resolution of the SAR, and G is the geometric factor given in equation (6). The spectral decrease over a limited wavenumber range, $\sigma_k < k_y < k_n$, can be approximated by a straight line on a log-log plot from which one can fit a power-law slope,

$$\langle S_c(k_x) \rangle = A k_x^{-\alpha} \quad (3)$$

where $\langle S_c(k_x) \rangle$ is the corrected power spectral density, α is the power-law spectral index, and the brackets $\langle \rangle$ indicate range averaging. The correction is required since the intensity spectrum includes an inherent wavenumber decrease due to the radar system impulse response. The impulse response was based on a least square fit to a Gaussian function using a processing area over land. The flat spectral noise floor for $k_x > k_n$ illustrates that the correction has been applied correctly. The spectral width (σ_k) and power-law index (α) were scaled from the intensity spectra.

IV. RESULTS

The analysis method described above allows estimates of the spectral width (σ_k) and spectral slope (α) for a variety of radar and ocean conditions during the NORCSEX experiment. In order to interpret the measured spectral shape, the relationship between the power spectrum of the ocean waveheight $\langle S_w(k_x) \rangle$ and the corrected intensity $\langle S_c(k_x) \rangle$ based on a linear transfer function [Ref.1] is given by,

$$\langle S_c(k_x) \rangle = \langle \sigma_0^2 \rangle |F(k_x)|^2 |R_b(k_x)|^2 \langle S_w(k_x) \rangle \quad (4)$$

where σ_0 is the radar cross section $F(k_x)$ is the velocity smearing filter given by equation (1), and $R_b(k_x)$ is the velocity bunching transfer function,

$$|R_b(k_x)|^2 = (R/V)^2 g k_x^3 G^2(\theta, \phi) \quad (5)$$

where ϕ is the azimuth angle, $\sin \phi = k_y/k$, θ is the angle of incidence, g is the acceleration of gravity and the geometric factor $G^2(\theta, \phi)$ is,

$$G^2(\theta, \phi) = \cos^2 \theta + \sin^2 \theta \sin^2 \phi. \quad (6)$$

The analysis was separated into two parts in order to study the variation of σ_k and α with radar and ocean parameters.

First, we selected an example when a single ocean wave system was traveling in the range direction, $\phi \approx \pi/2$. This criteria constrains the velocity bunching to a linear region. Three different processing areas within the swath were selected for analysis centered along a line perpendicular to the flight path with a wide variation in the R/V ratio. The R/V values were taken at the center of these areas, designated near, mid, and far range.

The spectral parameters for these three areas are given in Table I. The near and far range spectra were given in Fig. 1. The wave conditions were moderate with a southerly traveling wave system and a light wind from the south, $H_{1/3} \approx 2$ m, $U \approx 2$ m/s, and a dominant wavelength of about 180 m. c_{br} is a velocity bunching parameter describing the degree of the non-linearity [Ref. 8]. Linear velocity bunching is associated with $c_{br} \leq 0.3$. The March 14 pass 4 observations at near, mid, and far ranges are in the linear range based on the above criteria, see Table I. The σ_k and α values were calculated using a regression analysis (least square fit) between equation (1) and equation (3) and the corrected spectral data points. The correlation coefficient (R) is a measure of the goodness of fit. It has a range from 0 to 1 with 1 indicating a perfect fit. This regression analysis resulted in $R > 0.97$ for all cases. This indicates an good fit to both the Gaussian and power-law functions over the available wavenumber range.

The results given in Table I show several interesting features. First, the spectral index α varies with R/V such that a flatter spectra correspond to the larger

R/V values. The waveheight spectrum of the form $S_w(k_x) = B k_x^{-\beta}$ is assumed. This waveheight form implies that $\alpha = 3 - \beta$, based on equation (4) where $\beta \approx 3.5 - 4.0$ [Ref. 9]. Therefore the flat spectra $\alpha \approx 1.0$ at the far range (see Table I) tend to be consistent with a power-law slope based on linear velocity bunching. Secondly, the rms velocity σ_v shows a small decrease with increasing R/V. The σ_v values based on equation (2) are independent of R/V provided that, (1) the geometric factor $G(\theta, \phi)$ and the wave spectra do not change across the swath width, (2) the rms facet velocity controls the spectral shape, and (3) bunching non-linearities are negligible across the swath width. The geometric factor is $G \approx 1$ over the swath width since $\phi \approx \pi/2$, and the non-linearities due to velocity bunching are thought not to bias σ_v since $c_{par} < 0.3$.

The relative importance of three factors: the azimuth cutoff, non-linear velocity bunching, and the wave height spectrum in forming the spectral width is an area we confront. Velocity smearing causes a steeper azimuth spectral decrease than velocity bunching. This dependence is shown using eq. (4) where $|R_p|^2 \langle S_w \rangle \approx B k_x^{-1}$ (for $k_x > k_p \cos \phi$) changes slowly with k_x compared to $|F|^2$ since $|R_p|^2 \langle S_w \rangle$ reaches its maximum at $k_x = k_p \cos \phi$ and $\sigma_k \approx k_p \cos \phi$ where k_p is the peak wave number in the ocean wave spectrum. Indeed, a linear dependence of σ_x on R/V is one important indicator that the azimuth filtering is the dominant factor of the three. Therefore a linear regression analysis was performed on the three σ_x values of Table I resulting in slope estimate of $\sigma_v = 0.25$ m/s and $R = 0.82$ which includes the condition, $\sigma_x = 0$ at $R/V = 0$. The flatness of the power-law slope at the far range suggests that the waveheight spectrum may not be important in determining the observed spectral width whereas the correlation coefficient $R = .82$ suggests that the filtering is the dominant factor in determining the observed spectral width. Additional quantitative measurements similar to those given above are needed to further investigate the relative importance of factors responsible for the spectral width. In the present work, the velocity smearing is assumed to be the only factor influencing the spectral width when $c_{par} < 0.3$.

Second, we selected an ocean region where an atmospheric front was clearly visible. These fronts appear as bright and dark background areas in the SAR imagery. The March 17 pass 4 radar measurements at 0930 UT were selected because of the available surface truth across the front; namely, ship and buoy measurements of winds and waves. The front boundary was located along a North-South line with the higher winds to the west with the front moving from west to east. The surface wind changed from about 2 to 10 m/s across the front at about 1100 UT based on the ship measurements at a height of 14 m whereas the buoy was located about 60 km to the east of the SAR measurements. The buoy measurements show the front passing the buoy sometime between 1200 and 1500 UT since these measurements were taken at three hour intervals. The 1500 spectrum shows a wind driven spectrum with a wider angular spread and high frequency components for $H_{1/3} \approx 3.4$ m, whereas the 1200 spectra show a single peaked narrow spectrum (ocean wave swell) with $H_{1/3} \approx 2.0$ m and the high frequency component dissipated due to the decrease in wind velocity, as shown in Figure 2. The radar cross section change is approximately 10 dB. Figure 3 shows the spatial intensity variation or equivalently the radar cross section change across a scan perpendicular to the front. This radar scan, given in Fig. 3, was selected at the ships' intersection with the front, scans at the radar processing areas are within about 20 km of the ship measurements. They show a similar radar cross sectional variation. Fronts

are visible due to changes in the radar cross section which responds almost instantaneously to the local wind since the Bragg wavenumber is $k_B \approx 2\pi/5$ cm at C-band. The SAR processing areas were each selected approximately 5 km from the front boundary.

The images were processed using the same procedure as in the first example. It was not possible to obtain 512×512 processing areas with the same R/V ratios on both sides of the front and still maintain the criteria of range traveling ocean waves. Several additional problems were encountered in processing: (1) changes in the wave direction across the front can bias the spectral width unless care is taken in the analysis, and (2) the closeness of the processing area to the front boundary can bias the measurements. The front analyzed here is both an atmospheric front and an ocean front since the long waves propagating through the front change direction at the boundary. This change in direction is shown in the two-dimensional spectral plot of Figure 4. The spectral contour interval is 2 dB with a full range of 10 dB. The concentric circles correspond to wavelengths from 100 to 500 m. The change in direction is clearly shown with the waves traveling more nearly in range. The azimuth cutoff is most visible here and the high (ground-range) wavenumber spectral density more noticeable in the bright area (high wind region) than the dark region (low wind region).

The spectral parameters given in Table II for the atmospheric front show several interesting features. First, and most important, the observed spectral width shows little or no variation with surface winds. The wind varied between 2 and 10 m/s across the front. Indeed, a linear regression analysis using the same procedure as for the March 14 data was accomplished for the four σ_x values in Table II. The slope estimate is $\sigma_v = 0.30$ m/s and the goodness of fit is $R = 0.96$ resulting in a better linear dependence on R/V than the March 14 example.

The analysis shows that velocity smearing $\sigma_v \approx 0.30$ m/s is independent of the surface wind speed for the C-band SAR measurements based on an azimuth resolution of 10 m and $\sigma_x = (R/V) \sigma_v$. An estimate of σ_v is required for numerical modeling of moving ocean waves [Ref. 10]. The velocity smearing σ_v can be extrapolated to the resolution of the ERS-1 SAR under certain conditions. The resolution area was assumed equal in azimuth and range with a square root dependence on resolution [Ref. 5]. The resulting ERS-1 velocity smearing is $\sigma_v \approx 0.5$ m/s and the spatial smearing in $\sigma_x \approx 55$ m.

V. CONCLUSIONS

Special processing of SAR measurements has provided quantitative information on distortions in wave images due to ocean wave motion. The two mechanisms which cause these distortions can be quantified provided care is taken in selecting the radar/ocean conditions when these mechanisms are well understood. The physical processes underlying these distortions are (1) the spread in the radial facet velocities within a radar resolution cell, and (2) non-linear velocity bunching. The SAR measurements selected for analysis were restricted to cases when the ocean waves were traveling in or near the range direction. Range traveling ocean waves mean that velocity bunching is small. The azimuth spectra could then be analyzed quantitatively in terms of the remaining distortion mechanism, velocity smearing. This distortion is used to extract the rms radial facet velocities by measuring the spectral width for a range of radar/ocean parameters.

The linear dependence of spectral width with R/V is evidence that the velocity smearing is the only factor influencing the spectral width. The special processing and analysis shows that the spectral width is

independent of, or in the worst case only weakly dependent on, the surface wind speed. This inference was based on measurements across an atmospheric front where the wind from one side of the front to the other changed from 2 to 10 m/s. The rms facet velocities inferred from all the measurements in the present work varied between $\sigma_v \approx 0.25$ to 0.3 m/s. The buoy wave measurements are in agreement with the SAR measurements analyzed in the present work.

The relevant spatial scale which enters into the velocity smearing is the degraded radar resolution cell size rather than the stationary resolution given by others [Ref. 4-6]. The degraded cell size is appropriate because: First, this scale is consistent with the measurements reported here; namely, the short Bragg waves respond quickly to changes in the surface wind whereas the velocity smearing is independent of wind. This suggests the scale that contributes to the smearing is probably larger than the stationary resolution of 10 m, say 50 to 100 m. Indeed, this independence of wind implies a large resolution, since the short waves, the order of 10 m, are dissipated when the wind changed from 10 m/s to 2 m/s, (Fig 2). Second, the radar obtains its coherence and resolution by locating adjacent resolution cells relative to each other along the flight direction. Imaging radars are coherent in terms of locating adjacent cells only over a cell size given by the degraded resolution not the stationary resolution. Furthermore, the underlying physical process(es) and the details of how facet velocities enter into the azimuth smearing is not yet well understood. Additional work is needed to better understand how the facet velocity smearing enters into the SAR imaging of ocean waves.

ACKNOWLEDGEMENTS

We thank the Canadian Centre for Remote Sensing for furnishing the digital SAR data and Norman Malinas for processing the images. This work was supported by the Office of Oceanographer of the Navy, SPAWAR, and the Office of Naval Research (ONR) Contract #N00014-81-C-0692.

REFERENCES

- [1] Beal, R. C., D. G. Tilley, F. M. Monaldo, "Large- and Small-Scale Spatial Evolution of Digitally Processed Ocean Wave Spectra from SEASAT Synthetic Aperture Radar," J. Geophys. Res., vol. 88(c3), pp 1761-1777, 1983.
- [2] Vesecky, J. F., and R.H. Stewart, "The Observation of Ocean Surface Phenomena using Imagery from the SEASAT Synthetic Aperture Radar," J. Geophys. Res., vol. 87, pp 3397-3430, 1982.
- [3] Shuchman, R. A., J. D. Lyden, and D.R. Lyzenga, "Estimates of Ocean Wavelength and Direction From X-band and L-band Synthetic Aperture Radar during Marineland Experiment," IEEE J. Oceanic Eng. vol. Oe-8, pp 90-96, 1983.
- [4] Hasselmann, K., R. K. Raney, W. J. Plant, W. Alpers, R. A. Shuchman, D.R. Lyzenga, C. L. Rufenach, and M. J. Tucker, "Theory of Synthetic Aperture Radar Ocean Imaging: A MARSEN View," J. Geophys. Res., vol. 90(c3), pp 4659-4686, 1985.
- [5] Alpers, W. R., and C. Bruening, "On the Relative Importance of Motion-Related Contributions to the SAR Imaging Mechanism of Ocean Surface Waves", IEEE Trans. Geosci. Remote Sens., vol GE-24(6), pp 873-885, 1986.
- [6] Tucker, M. J., "The Imaging of Waves by Satelliteborne Synthetic Aperture Radar: the Effects of Sea-Surface Motion", Int. J. Remote Sensing, vol. 6(7), pp 1059-1074, 1985.
- [7] Livingstone, C. E., A. L. Gray, and R. K. Hawkins, "CCRS C-band Airborne Radar - System Description and Test Results", 11th Canadian Symposium on Remote Sensing, Waterloo, Ontario, June 22-25, 1987.
- [8] Alpers, W. R., "Monte Carlo Simulations for Studying the Relationship Between Ocean Wave and Synthetic Aperture Radar Image Spectra", J. Geophys. Res., vol. 88(C3), pp 1745-1759, 1983.
- [9] Donelan, M. A., and W. J. Pierson, Jr., "Radar Scattering and Equilibrium Ranges in Wind-Generated Waves With Application to Scatterometry", J. Geophys. Res., vol. 92(C5), pp 4971-5029, 1987.
- [10] Lyzenga, D. R., "Numerical Simulation of Synthetic Aperture Radar Image Spectra for Ocean Waves", IEEE Trans. Geosci. Remote Sens., vol. GE-24 (6), pp 863-871, 1986.

TABLE I. NORCSEX Radar Parameters Near, Mid, and Far Ranges
14 March 1988 Pass 4

| Range | Incidence Angle, θ | Azimuth* Angle, ϕ | R/V (Sec) | α P-L Slope | σ_k (rad/m) | σ_x (m/rad) | σ_y (m/s) | C_{bar} |
|-------|---------------------------|------------------------|-----------|--------------------|--------------------|--------------------|------------------|-----------|
| Near | 65 | 84 | 113 | 2.7 | 0.027 | 37 | 0.32 | 0.21 |
| Mid | 76 | 84 | 196 | 1.7 | 0.019 | 52 | 0.27 | 0.23 |
| Far | 80 | 84 | 286 | 0.9 | 0.016 | 62 | 0.23 | 0.31 |

*Values based on SAR spectral peak direction

TABLE II. NORCSEX Radar Parameters across Atmospheric front
17 March 1988 Pass 4

Processed Data Area 1

| Wind | Incidence Angle, θ | Azimuth* Angle, ϕ | R/V (Sec) | α P-L Slope | σ_k (rad/m) | σ_x (m/rad) | σ_y (m/s) | C_{bar} |
|------|---------------------------|------------------------|-----------|--------------------|--------------------|--------------------|------------------|-----------|
| High | 72 | -83 | 168 | 0.55 | 0.021 | 48 | 0.28 | 0.19 |
| Low | 59 | 58 | 100 | 1.32 | 0.030 | 33 | 0.30 | 0.25 |

Processed Data Area 2

| | | | | | | | | |
|------|----|-----|-----|------|-------|----|------|------|
| High | 73 | -60 | 184 | 2.10 | 0.018 | 56 | 0.30 | 0.35 |
| Low | 64 | 57 | 120 | 2.60 | 0.026 | 38 | 0.32 | 0.28 |

*Values based on SAR spectral peak direction

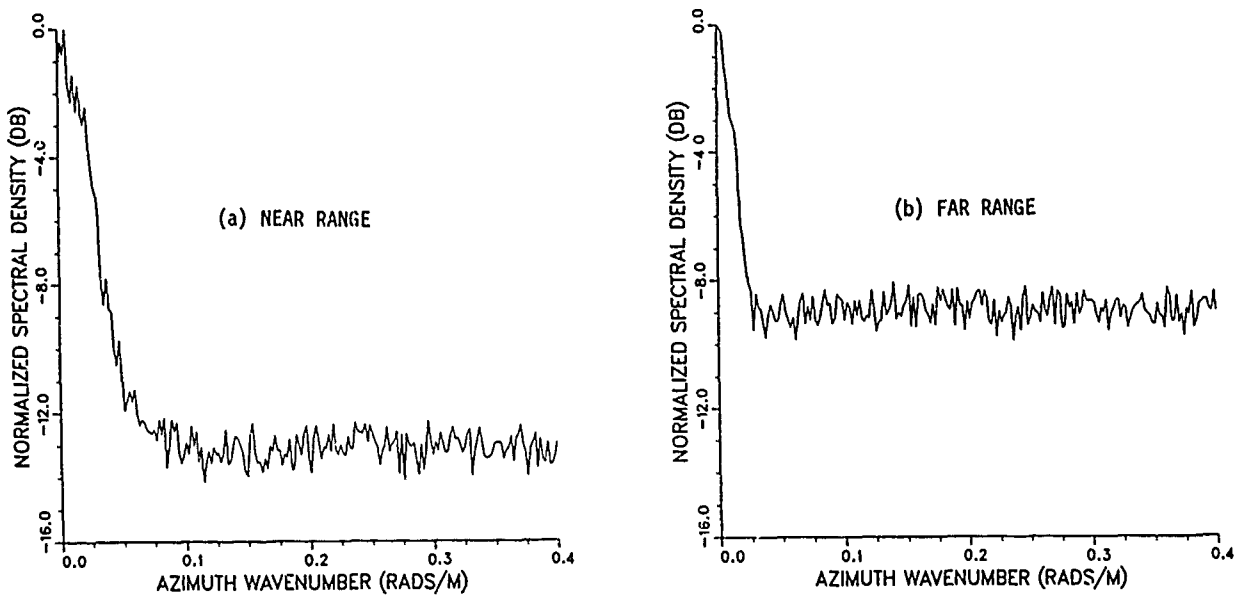


Figure 1. One-Dimensional Spectral Density for Near (a) and Far (b) edge of the Radar Swath Width

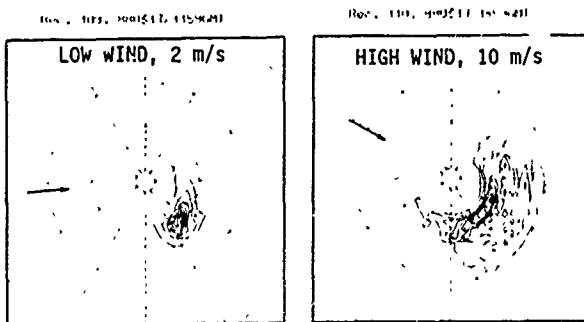


Figure 2.

Two-Dimensional P/R Buoy Spectra on 17 March at 1200 UT (low wind) and 1500 UT (high wind) acquired at 64.1° North, 7.3° East about 60 km East of the SAR Observation taken at 0930 UT. The Arrow Indicates Wind Direction and the Concentric Circles Represent Wave Frequencies at 0.05, 0.1, 0.2, and 0.6 Hz. The Winds Changed from U (3.7) = 3.3 m/s at 1200 to 11.3 m/s at 1500 While the Dominant Wave Period was ≈ 11 Seconds at Both Times

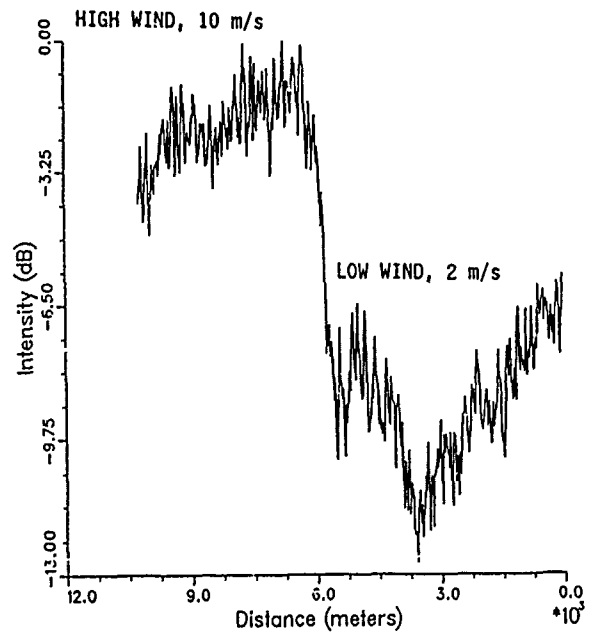


Figure 3. One-Dimensional Intensity Variation Across the Atmospheric Front on 17 March 1988

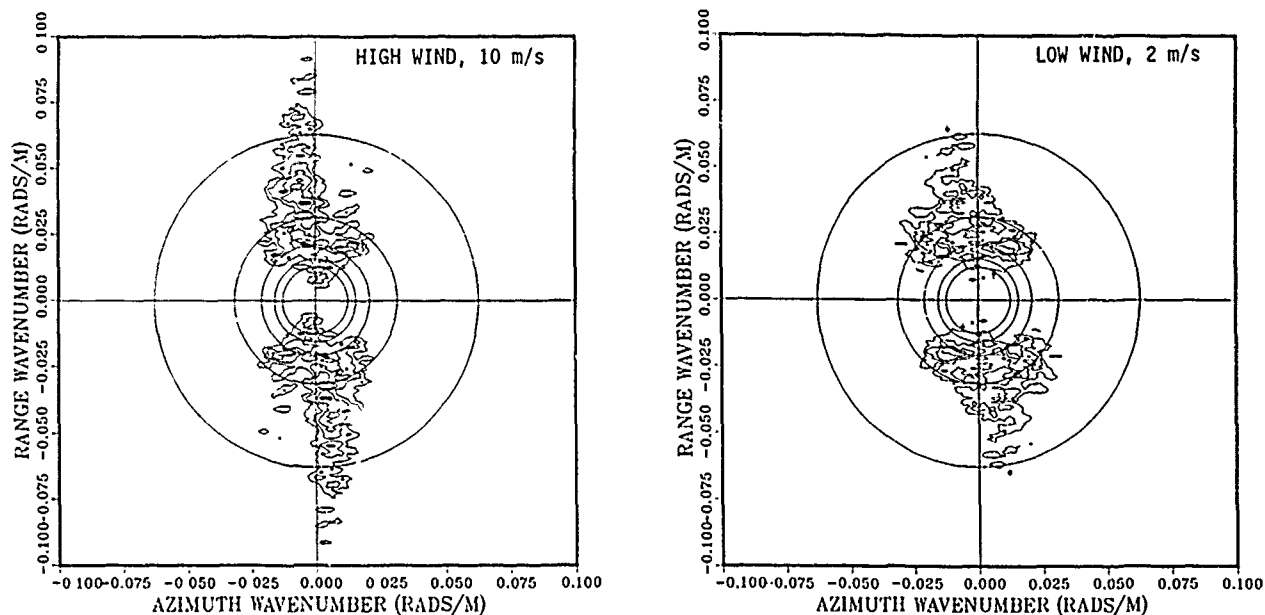


Figure 4.

Two-Dimensional SAR Spectral Contour Plots on
 Two sides of Atmospheric Front on 17 March
 1988 taken at 0930 UT (data area 1).
 Contour Levels are 2 dB and the Concentric
 Circles Correspond to 100 m intervals in
 Wavelength Varying Between 100 and 500 m

WAVE MEASUREMENTS ON HALTENBANKEN DURING NORCSEX'88:

An Intercomparison of Buoy, SAR and Altimeter Data

Richard B. Olsen, Stephen F. Barstow and Harald Schyberg

Oceanographic Company of Norway

Pirsenteret

N-7005 Trondheim

NORWAY

ABSTRACT

Results of an analysis of simultaneous measurements of ocean waves from directional wave buoys, the Canada Centre for Remote Sensing airborne SAR and the GEOSAT radar altimeter during the NORCSEX'88 ERS-1 pre-launch experiment on Haltenbanken in March 1988 are presented. The method suggested by Monaldo and Lyzenga (1986) for estimating wave height spectra from SAR images has been tested on the airborne SAR data for a case with a complicated wave field. The significant wave height estimates derived from the SAR spectra were compared with wave height estimates from the buoy and the altimeter. Comparisons of the directional spectra computed from the buoy and the SAR data are also presented.

Keywords: SAR, Directional Wave Spectra, Significant Wave Height, NORCSEX'88, WAVESCAN, GEOSAT, Altimeter.

INTRODUCTION

The Norwegian Continental Shelf Experiment (NORCSEX '88) took place on Haltenbanken, off the coast of Norway during March/April 1988. The experiment was carried out to gain increased knowledge about the use of microwave remote sensing instruments, especially Synthetic Aperture Radar (SAR), for monitoring the marine environment, as part of Norwegian preparations for the ERS-1 mission.

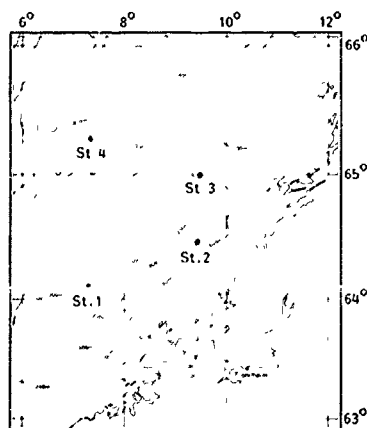


Fig. 1 The location of the four directional wave measurement buoys in the Haltenbanken area during NORCSEX'88.

During the period March 11th to 21st, six flights with the Canada Centre for Remote Sensing C/X-band SAR were carried out in the Haltenbanken region. The flights were dedicated to different investigations, depending on the wind/wave conditions encountered for each flight. These were monitored by the R/V "Håkon Mosby" from the University of Bergen, and 4 directional wave measuring buoys deployed by OCEANOR. Also, GEOSAT altimeter data were acquired in near real time for the duration of the experiment.

On two flights, March 11th and 20th, conditions were suitable for wave studies. This paper presents some of the SAR measurements from March 11th, and the entire NORCSEX GEOSAT data set. These are compared with relevant buoy observations. Further details may be found in [1] and [2].

BUOY MEASUREMENTS

Four directional wave metocean data buoys were moored in the Haltenbanken area during NORCSEX'88. Fig. 1 shows the locations of the buoys which were situated at cross-over points of the GEOSAT ground tracks. A WAVESCAN buoy was located at Station 1 whilst NORWAVE buoys were positioned at the other stations. Each buoy measured the directional wave spectrum in addition to wind speed, direction, air and sea temperature and air pressure. Data were stored on magnetic tapes on board and a data summary consisting of wave and meteorological parameters was transmitted by satellite to land in real time. Both buoy systems were extensively tested during extreme and moderate sea states in WADIC (Wave Direction Calibration Project) and comprehensive comparisons of directional wave spectra and parameters are available referenced to an array of instrumentation on a fixed offshore platform (see ref. 3).

The analysis of the buoy data is described in [2]. Time series of significant wave height, H_{m0} , and peak wave period, T_p , are shown in Fig. 2. A good range of wave heights was experienced; several storm events can be identified and, in addition, long period swells dominated during 23rd-25th March. Strong spatial variability in the wave field in the area is also apparent, particularly during the storm on 19th March, as a result primarily of large variations in fetch between stations for certain wave directions. For much of the experiment wave conditions were complex with crossing wave systems particularly during the SAR comparison period on 11th March when trimodal conditions occurred.

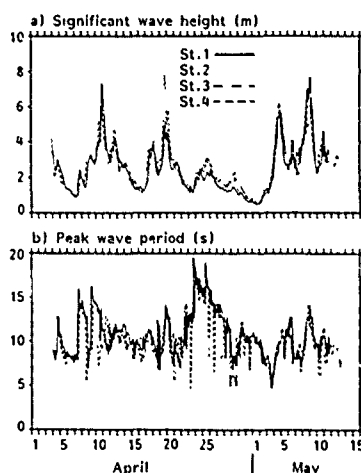


Fig. 2 Time series of a) significant wave height, H_m0 , and b) peak wave period, T_p , during NORCSEX'88.

GEOSAT MEASUREMENTS

The GEOSAT altimeter estimates of significant wave height were derived by NOAA from computations on board the satellite every second. These values were subsequently corrected for satellite attitude and averaged to give 10-second mean values. The wind speed data were computed using the Brown algorithm [4]. An example of data from one pass during NORCSEX'88 is shown in Fig. 3. The length of the line above the ground track is proportional to the wave height and similarly below the line to the wind speed.

A comparison of GEOSAT and buoy derived estimates of significant wave height, H_m0 , and wind speed (referenced to 10 m) is shown in Figs. 4 and 5. The criteria used for simultaneous data was that the distance from the ground track to the buoy is less than 50 km and that the two estimates are separated in time by less than 1.5 hours. The buoy estimate nearest in time to the GEOSAT measurement is used except for the pass at the storm peak on 11th March. Here, as the GEOSAT pass occurred approximately midway between two buoy measurements and as conditions were strongly non-stationary we have used a linearly interpolated buoy estimate (see Fig. 7).

All the GEOSAT estimates on each pass satisfying the simultaneity criteria are included on the plots so that we also get an idea of the spatial variability along the track.

We note that for H_m0 there is a trend for GEOSAT to underestimate wave heights particularly for the higher sea states. The largest outliers are mostly measurements made close to the coast, and if we omit these points the mean difference is 0.45 m with rms difference of 0.48 m. Similar results were found in [5] from a larger data set based on the NDBC buoy network. A mean difference of 0.36 m and a rms difference of 0.49 m was found in that study. Uncertainty is highest for the high sea states as there is few data available. It was found in [3] that several wave buoys underestimate significant wave height in high seas. The accuracy of the NDBC buoys in high sea states would not seem to be well known. A comparison relative to one of the buoys tested in [3] would certainly be valuable in comparing results from the US with the work now being undertaken in Norway.

The wind speed comparison (Fig. 5) shows a lot of scatter, with a mean difference of 0.8 m/s and rms difference of 3.5. The tendency for GEOSAT to overestimate at low speeds and to underestimate at high speeds found in [5] is also apparent in the present data set. Further work on the accuracy of the buoy wind measurements is ongoing.

SAR MEASUREMENTS

On March 11th, seven flight passes with the CCRS C/X SAR system were made over the moored directional wave buoy at station 1. Four passes were at approx. 20,000 ft, while three were repeated at approx. 12,000 ft. The SAR imagery was acquired in Nadir Mode (see [6] for a detailed system description). The lines were designed to obtain azimuth travelling and range travelling waves, with additional passes imaging waves travelling at 45° relative to the aircraft ground track. As may be seen in Fig. 6, the sea state was very complex with three dominant peaks indicating waves propagating towards south, southeast and east. We did not, therefore, obtain flight lines with purely azimuth or range travelling waves. However, for most flight lines the two-dimensional image spectra (Fig. 6) indicate both near range and near azimuth travelling waves.

In order to compare directional spectra estimates from the directional wave buoy with SAR image spectra, the procedure outlined in [2] was followed. Imagery centred on the buoy position was slant-to-ground range corrected before further analysis. In order to acquire reliable spectral estimates, 15 FFTs from blocks of 512x512 pixels were computed, each taken at the same ground range, and subsequently averaged to form one image spectral estimate. As the airborne SAR is a relatively slow moving line scanner, the images and their spectra suffer from so-called scanning distortion. As the ambiguity in the image spectra leads to some problems in correcting for this, it is easier to distort the buoy spectrum in order to compare with the SAR image spectrum (see [7]). In our comparisons the buoy spectra were made symmetric and ambiguous, to simulate an "idealized image spectrum". In order to obtain spectra acquired at incidence angles similar to previous experiments (SEASAT, SIR-B), the sub-images were extracted in the 20-40° incidence angle range, with an R/V ratio of 65 s and 28 s for the high and low altitude passes respectively. The SAR image spec-

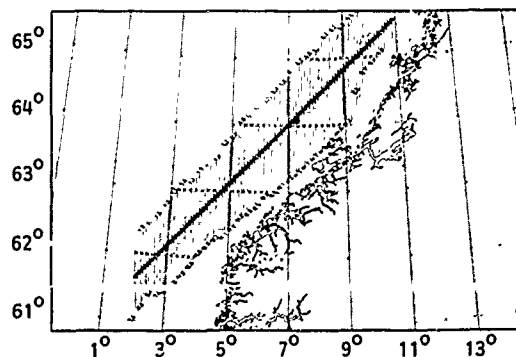


Fig. 3 GEOSAT altimeter ground track early on 11th March. Vertical lines represent altimeter significant wave height estimates (up) and for wind estimates (down). The location of station 1 is also shown.

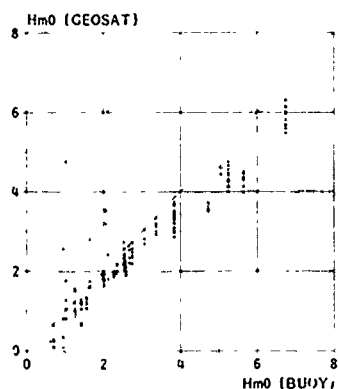


Fig. 4 Scatter plots of significant wave height (H_{m0}) from the GEOSAT altimeter vs. in-situ buoy estimates.

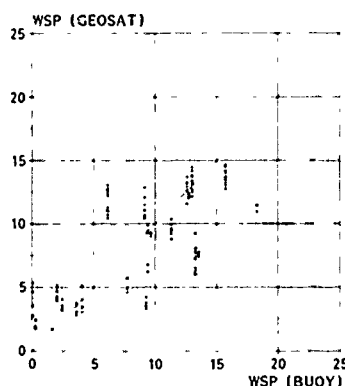


Fig. 5 Scatter plot of wind speed estimate (m/s) (Brown algorithm) from GEOSAT vs. in-situ buoy measurements.

tra and corresponding scan-distorted WAVESCAN spectra are shown in Fig. 6. Circles indicate 100, 200 and 400 m (inner circle) wavelengths. The SAR spectra are compared with the buoy spectra measured closest in time to the overflights, 12.00 GMT for the high altitude passes and 15.00 GMT for the low altitude passes. At present, only visual comparisons have been made. The most dominant features of the WAVESCAN spectra are generally well represented in the SAR spectra, although to varying degrees in the different passes. There is definitely an R/V dependence in the azimuth wave number response, although no comparisons with existing models have been made as yet.

APPLICATION OF SAR TRANSFER FUNCTIONS

Various models describing the response of a SAR to an ocean surface wave field are available, see e.g. [8] to [11]. A practical method for obtaining "calibrated" wave height spectra by estimating speckle noise variance and applying a Modulation Transfer Function (MTF) correction based on a model similar to that of [9], has been outlined in [12] (in the following referred to as the APL procedure). If this were a realistic approach, a significant wave height estimate, defined as $\sqrt{m_0}$, where m_0 is the zeroth order moment of the spectrum, would be obtainable.

Few such wave height estimates have been reported in the literature. As we here have a complicated sea state case with near simultaneous observations from a WAVESCAN buoy, the GEOSAT alti-

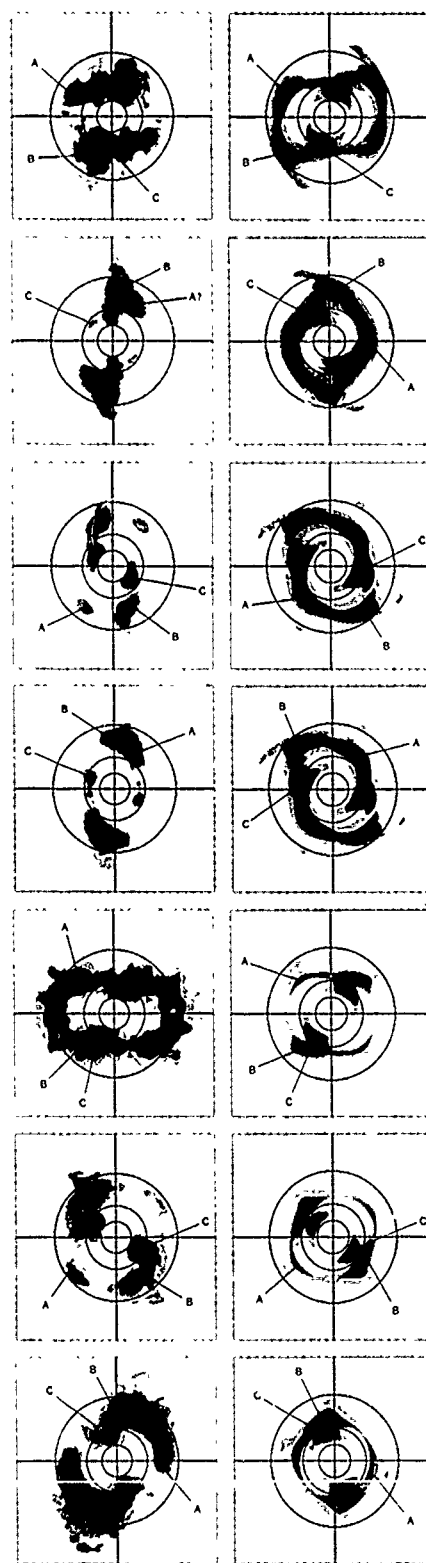


Fig. 6 SAR (left) and WAVESCAN directional spectra, lines 1-7, 11 March 1988. Circles indicate 100 m, 200 m and 400 m wave lengths. Horizontal axis is azimuth and vertical is range dimension.

meter and an airborne SAR, it is of interest to employ the APL procedure to see if the SAR derived wave height is realistic and comparable with the other measurements. As the APL procedure is based on several fundamental assumptions, leading to a theoretical, linearized version of the MTF, care must be taken when analyzing airborne SAR imagery in order not to exceed the validity of the linearizations. The image data in this analysis were extracted at 20-40° incidence angle, comparable to SEASAT and SIR-B, for which the MTF formulation was derived. The APL procedure was applied to the averaged spectra, and the results are presented in Table 1 and Fig. 7. Table 1 shows that the results were mostly comparable to the 4.2 m and 3.6 m significant wave height estimates obtained by the WAVESCAN buoy, although the SAR generally overestimates the wave height by a significant amount. A further comparison with the buoy and GEOSAT wave heights (Fig. 7) shows that the difference between the SAR and the buoy is not significantly greater than the difference between the buoy and GEOSAT, although altimeter wave heights are generally accepted as reasonably accurate.

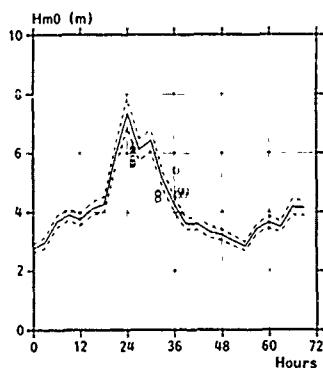


Fig. 7 Time series of WAVESCAN significant wave height with 90% confidence limits, 10-12 March 1988. Circles and crosses are GEOSAT observations, while squares are SAR estimates.

CONCLUSIONS

Comparison between buoy and SAR measurements show that in this case of a complicated sea state, the SAR obtained promising results.

Comparison of GEOSAT altimeter estimates of wave height with WAVESCAN and NORWAVE buoy data is reasonably good although with a tendency to increased overestimation at high sea states. The wind speed comparison shows more scatter but, along with the wave height results, is fairly consistent with results from earlier studies.

Table 1 Significant wave height estimates from the airborne SAR passes, 11.MAR.88.

| Line | Hm0 [m] |
|------|---------|
| 1 | 5.4 |
| 2 | 4.6 |
| 3 | 3.9 |
| 4 | 4.7 |
| 5 | 0.1 |
| 6 | 4.6 |
| 7 | 4.7 |

ACKNOWLEDGEMENTS

This work was funded by Operator Committee North (OKN) and the Norwegian Space Center. We gratefully acknowledge their support.

LITERATURE

- [1] OLSEN, R.B. and S.F. Barstow, 1988: Wave Measurements on Haltenbanken During NORCSEX '88; An Intercomparison of Buoy, Synthetic Aperture Radar and Altimeter Data. Oceanor Report no. OCN88081, Trondheim, Norway, 134 pp.
- [2] BARSTOW, S.F. and S. Bjerken, 1988: Wave Measurements from Moored Directional Wave Buoys During NORCSEX '88. Oceanographic Data Acquisition Project. Report no. 108, Trondheim, Norway, 146 pp.
- [3] ALLENDER, J., T. Audunson, S.F. Barstow, S. Bjerken, H.E. Krogstad, P. Steinbakke, L. Vartdal, L. Borgman and C. Graham, 1989: The WADIC project: A Comprehensive Field Evaluation of Directional Wave Instrumentation. Submitted to Ocean Engineering.
- [4] BROWN, G.S., 1979: Estimation of Surface Winds Using Satellite-Borne Radar Measurements of Normal Incidence. J. Geophys. Res. Vol. 84, 3974-78.
- [5] DOBSON, E., F. Monaldo, J. Goldhirsh and J. Wilkerson, 1987: Validation of GEOSAT Altimeter-Derived Wind Speeds and Significant Wave Heights Using Buoy Data. John Hopkins APL Techn. Dig. Vol. 8 (2), 222-33.
- [6] LIVINGSTONE, C.E., A.L. Gray, R.K. Hawkins, J.G. Halbertsma, R.A. Deane and R.B. Olsen, 1987: CCRS C-band Airborne Radar - System Description and Test Results. Proc. 11th Canadian Symp. on Remote Sensing, Waterloo, Ontario, June 22-25 1987.
- [7] KROGSTAD, H.E. and R.B. Olsen, 1988: An Intercomparison of SAR and Buoy Directional Wave Spectra from LEWEX. Proc. IGARSS '88 Vol. I, pp. 147-148. Edinburgh, Scotland 12-16 September 1988.
- [8] ALPERS, W.R., D.B. Ross and C.L. Rufenach, 1981: On the Detectability of Ocean Surface Waves by Real and Synthetic Aperture Radar. J. Geophys. Res. Vol. 86, pp. 6481-6498.
- [9] HASSELMANN, K., R.K. Raney, W.J. Plant, W. Alpers, R.A. Shuchman, D.R. Lyzenga, C.L. Rufenach and M.J. Tucker, 1985: Theory of Synthetic Aperture Radar Ocean Imaging: A MARSEN View. J. Geophys. Res. Vol. 90, pp. 4659-4686.
- [10] RANEY, R.K., 1981: Wave Orbital Velocity, Fade and SAR Response to Azimuth Waves. IEEE J. Oceanic Eng. Vol. OE-6, No. 4, pp. 140-146.
- [11] KASILINGHAM, D.P. and O.H. Shemdin, 1988: Theory for Synthetic Aperture Radar Imaging of the Ocean Surface: With Application to the Tower Ocean Wave and Radar Dependence Experiment on Focus, Resolution and Wave Height Spectra. J. Geophys. Res. Vol. 93, pp. 13837-13848.
- [12] MONALDO, F.M. and D.R. Lyzenga, 1986: On the Estimation of Wave Slope- and Height-Variance Spectra from SAR Imagery. IEEE Trans. Geoscience and Remote Sensing Vol. GE-24, pp. 543-551.

EXPERIMENTAL INVESTIGATION OF IMAGING GEOMETRY EFFECTS ON SAR-DERIVED OCEAN WAVE SPECTRA

K.A. Høgda, J.P. Pedersen, T. Guneriussen, H. Johnsen,

T. Eltoft,

FORUT,
P.o. Box 2806 Elverhøy,
N-9001 TROMSØ,
NORWAY

UNIVERSITY OF TROMSØ
P.o. Box 953,
N-9001 TROMSØ,
NORWAY

ABSTRACT

Along track (azimuth) filtering of synthetic aperture radar (SAR) ocean image spectra has been investigated under different range to velocity (R/V) ratios. Cut-off wavelength in azimuth is estimated from SAR ocean image spectra obtained with different R/V ratio. For a R/V=130 s which corresponds to the ERS-1 value, the cut-off azimuth wavelength obtained is $\lambda_{\text{cut}}=240$ m. A coherence time is estimated to be $\tau_{\text{coh}}=0.015$ s. The azimuth filtering is shown to influence both the directivity and the peak wavelength of the ocean image spectra. The incidence angle and thus also the integration time has negligible effect on the azimuth filtering over the range of angles and integration times considered.

Keywords: Synthetic Aperture Radar, Ocean Wavenumber Spectrum, Coherence Time.

1. INTRODUCTION

One of the major limitations in SAR imaging of ocean waves is the loss of resolution in the azimuth direction caused by the non-stationarity of the scattering ocean surface. The fine resolution in the azimuth direction, obtained by recording the phase histories of the backscattered signals over a finite time interval, consists of both constructive and destructive effects depending strongly on the radar and ocean wave parameters involved (Hasselmann et al., 1985). The constructive effect is theoretically described by the velocity bunching modulation mechanism and the destructive effect is usually modeled by a non-stationary response function. Of particular importance for both of these mechanisms is the R/V ratio and the sea-state condition being imaged (Beal et al., 1983), (Monaldo et al., 1983).

One of the objectives of the wave imaging experiment during the NORCSEX '88 campaign was to investigate SAR imaging of complex sea-states under various height to velocity ratios. Data from the march 11, are used in the present analysis, providing star-pattern flights at two heights and a trimodal wave spectrum with a significant waveheight of 4 m.

Results from investigation of the resolution loss in azimuth with respect to the R/V ratio and the angle of incidence are presented.

2. OCEAN IMAGE SPECTRA AND NON-STATIONARY RESPONSE FUNCTION

A linear relationship between SAR image intensity and the ocean surface height displacement yields a simple relation between the image power spectrum, $F(k)$, and the directional ocean wave spectrum, $W(k)$ (Beal et al., 1983):

$$F(k) = W(k) |M(k)|^2 |H(k)|^2 |G(k)|^2 \quad (1)$$

where $M(k)$ is the modulation transfer function, usually splitted into tilt-, hydrodynamic- and velocity bunching modulation. $H(k)$ is the stationary SAR impulse response function i.e. the SAR system response to a stationary point-scatterer. Finally, $G(k)$, is the non-stationary response function describing the SAR system response to a moving scatterer.

The non-stationary response function in eq.(1) models the loss of azimuth resolution as a low-pass filtering process on the image spectrum. Both acceleration and velocity spread effects may contribute to the response function (Beal et al., 1983), (Hasselmann et al., 1985). The first is proportional to the R/V ratio and the SAR integration time and the second is proportional to the R/V ratio and the scatterer velocity variance within a resolution element. The velocity spread contribution, being the dominant one, arises from the well known azimuth shift $\Delta y = RU_r/V$ introduced by the scatterer velocity U_r in the radar look direction. Assuming the velocity spread (i.e. also Δy) between different scatterer within a resolution element to be Gaussian the response function $G(k)$ also becomes Gaussian:

$$G(k) = \exp \left[-\frac{k_a^2}{2\delta k_a^2} \right] \quad (2)$$

where $\delta k_a = 1/\sqrt{2\langle\Delta y^2\rangle}$ is the half-width and $\sqrt{\langle\Delta y^2\rangle}$ is the rms. of the random position shifts. The half-width can thus be written as:

$$\delta k_a = \frac{V}{R\sqrt{2\langle U_r^2\rangle}} \quad (3)$$

Attempts to estimate δk_a (i.e. $\langle U_r^2\rangle$) or the corresponding cut-off wavenumber defined as $k_{\text{max}} = 2\delta k_a$ are in the literature done integrating a theoretical wave spectrum modelling the imaged sea-state (Tucker, 1985), (Monaldo et al., 1986). The velocity spread contribution may also alternatively be expressed by a coherence time

$$\tau_{\text{coh}} = \frac{\lambda_{\text{radar}}}{2\pi\sqrt{2\langle U_r^2\rangle}} \quad (4)$$

that effectively limits the available integration time and thus reduces the resolution in the azimuth direction. In this paper azimuth cut-off, k_{max} , and coherence time, τ_{coh} , are estimated directly from the SAR ocean image spectra. This is done utilizing spectra obtained under various R/V ratios using the procedure described in the following. According to eq. (1) the filtered azimuth image spectral profile $F(k_a)$ may be expressed by the corresponding unfiltered profile, $F(k_a)$ as:

$$F(k_a) = F(k_a)/G(k)^2 \quad (5)$$

Dividing spectral profiles F obtained with different R/V ratios but otherwise under same wave/imaging geometry yields then using eq. (2) and eq.(3):

$$\frac{F_2(k_a)}{F_1(k_a)} = \exp\left[-\frac{k_a^2}{2\delta k_{a2}^2} \left(1 - \frac{R_1^2 V_2^2}{R_2^2 V_1^2}\right)\right] \quad (6)$$

The azimuth half-width δk_{a2} may thus be estimated from the half-width of the function in eq. (6). Having established δk_{a2} , the velocity variance $\langle U_r^2 \rangle$ may be obtained from eq. (3) and finally the coherence time from eq. (4).

3. SAR DATA PROCESSING

Data from March 11. Line 3, 4 and 6 have been analyzed in this paper. The aircraft track angle, velocity and height varied as given in table 1.

| | Track angle | Velocity | Height |
|--------|-------------|-----------|--------|
| Line 3 | 59.4° | 107.3 m/s | 6126 m |
| Line 4 | 240.4° | 108.3 m/s | 6096 m |
| Line 6 | 59.5° | 149.3 m/s | 3688 m |

Table 1. Aircraft track angle, velocity and height.

The real time CV-580 SAR images are preprocessed and converted to ground range given an out pixel spacing of 7.7 m in both range and azimuth.

Out of this converted images, four subimages with 256 pixels x 1024 lines are selected at different ground ranges (and thus also incidence angles and integration times) as given in table 2.

| Sub Image | 1 | 2 | 3 | 4 |
|-----------|----------------|----------------|----------------|----------------|
| | R/V θ T | R/V θ T | R/V θ T | R/V θ T |
| Line 3,4 | 70 38° 0.37s | 95 55° 0.50s | 130 65° 0.68s | 170 71° 0.89s |
| Line 6 | 30 38° 0.16s | 40 55° 0.22s | 55 65° 0.30s | 70 71° 0.38s |

Table 2. R/V , incidence angle θ , and effective integration time T per look of the different subimages.

The images show a strongly varying brightness from near to far range. For each subimage an illumination function is computed by averaging all the range lines and fitting a third degree polynomial. All the range lines are then divided by this polynomial. The subimage power spectra are computed using the periodogram method and subsequently smoothed by a Gaussian kernel and filtered by an exponential high pass filter.

The spectra are corrected for the stationary SAR system response function, but, no correction is done for the modulation transfer function. This is not critical for the present analysis. The azimuth profiles used in the computations are in addition corrected for the along track scale distortion (Vachon et. al., 1988).

The power spectra are transformed into polar coordinates, $D(k, \theta)$, using bilinear interpolation. The directivity over all wavenumbers and the heave spectra are then obtained by integrating $D(k, \theta)$ with respect to wavenumber and angle, respectively.

4. EXPERIMENTAL RESULTS

Wavescan buoy derived ocean wave spectrum obtained during the SAR overflights is shown in fig 1 and the three main modes are numbered 1, 2 and 3. The direction of the SAR flight lines 3, 4 and 6 are also annotated on to the wave spectrum together with the local wind direction. The significant waveheight was between 6-7 m during the time of the SAR overflights. The wavescan spectrum below was used for identification of the various modes in the corresponding SAR ocean image spectra.

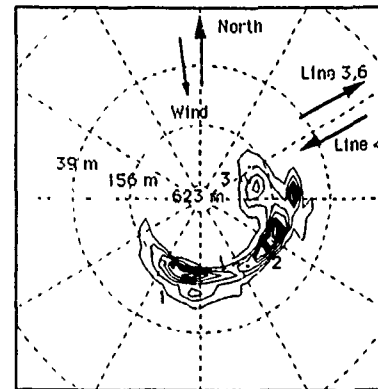


Fig. 1 Wavescan buoy derived ocean wave spectrum at the time of the SAR overflights. Local wind direction and SAR Line 3, 4 and 6 indicated by arrows. Significant waveheight about 6-7 meters. Contour plot taken from ODAP report no. 108 (Barstow et. al., 1988).

In fig. 2, 3 and 4 SAR ocean image spectra obtained near nadir (fig. a) and far nadir (fig. b) from Line 3, Line 4 and Line 6 are presented. The direction of flight is along the positive azimuth axis. The three modes indicated in the buoy spectrum (fig. 1) are marked with the same numbers in the SAR spectra.

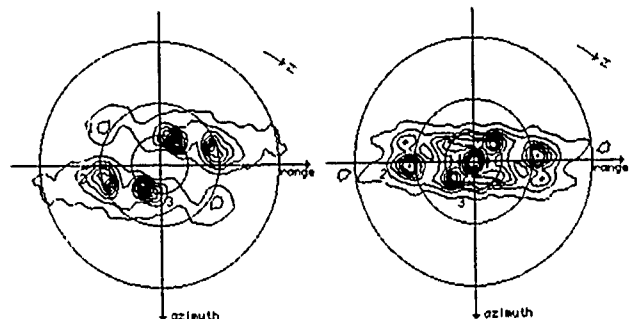


Fig. 2. Line 3 near and far nadir ocean image spectra. Outer contour 75 m, mid contour 150 m, inner contour 300 m. N indicates North.
a: $R/V=70$ s, incidence angle $\theta=38^\circ$.
b: $R/V=170$ s, incidence angle $\theta=71^\circ$.

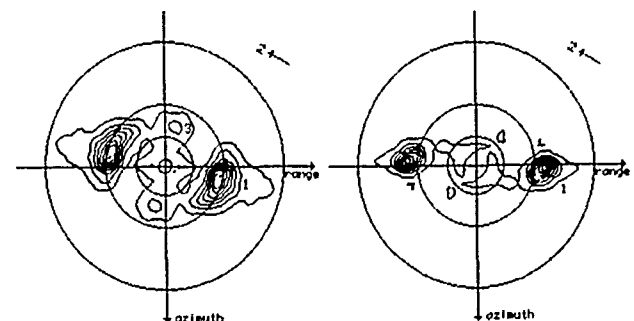


Fig. 3. Line 4 near and far nadir ocean image spectra.
a: $R/V=70$ s, incidence angle $\theta=38^\circ$.
b: $R/V=170$ s, incidence angle $\theta=71^\circ$.

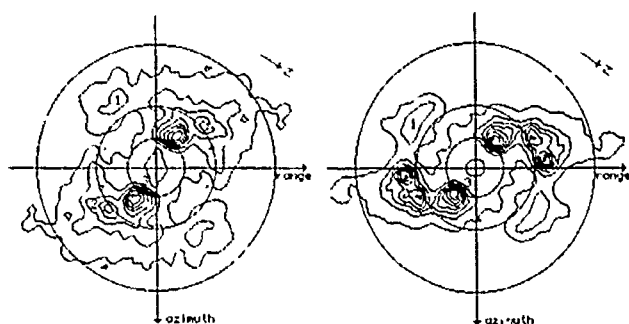


Fig. 4. Line 6 near and far nadir ocean image spectra.

a: $R/V=30$ s, incidence angle $\theta=38^\circ$.

b: $R/V=70$ s, incidence angle $\theta=71^\circ$.

The dominant effect demonstrated in these spectra is the increased azimuth filtering with increasing R/V ratio. Fig. 2a and fig. 4b which are taken under approximately same R/V ratio but very different incidence angle show close similarity. Fig. 2 and 3 shows spectra obtained at opposite directions of flight, and the modes with positive azimuth components are best identified. This will be a subject for further studies.

In order to investigate properly the azimuth filtering process indicated by the spectra in fig. 2, fig. 3 and fig. 4, azimuth profile at spectral peak and average azimuth profiles are computed for various R/V ratios and incidence angles. The filtering effect on the directivity and the heave spectrum is also considered.

Fig. 5a shows the azimuth profile averaged over all range lines in the spectrum obtained from Line 3 and Line 6 under different R/V ratios but same incidence angle. Fig. 5b shows the average azimuth profile obtained under different incidence angles but same R/V ratio.

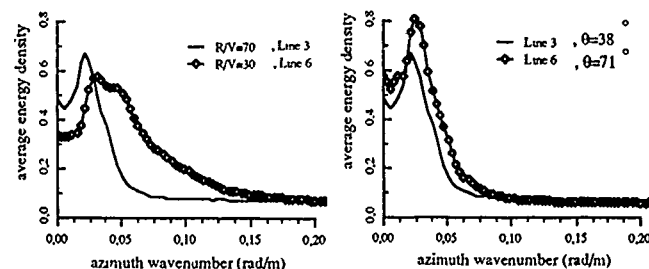


Fig. 5a. Average azimuth profile obtained from Line 3 and Line 6 under different R/V ratios but same incidence angle $\theta=38^\circ$.

b. Average azimuth profile obtained from Line 3 and Line 6 under different incidence angles but same $R/V=70$ s ratio.

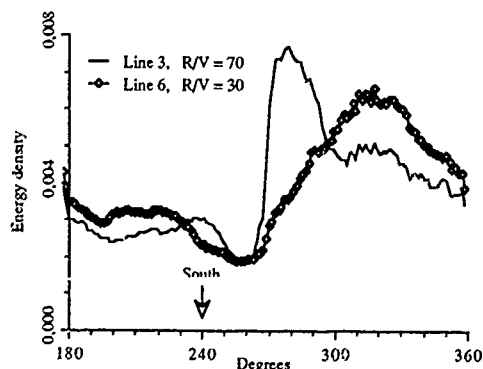


Fig. 6a Directivity obtained from Line 3 and Line 6 under different R/V ratios but same incidence angle $\theta=38^\circ$. Range corresponds to 270° .

The directivity and heave spectrum obtained under the same R/V ratios and incidence angles as in fig. 5 are shown in fig. 6 and 7, respectively.

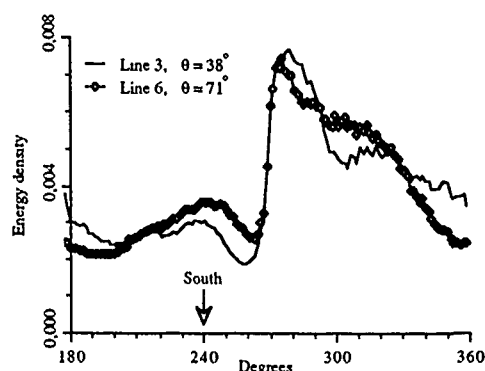


Fig. 6b Directivity obtained from Line 3 and Line 6 under different incidence angles but same $R/V=70$ ratio. Range corresponds to 270° .

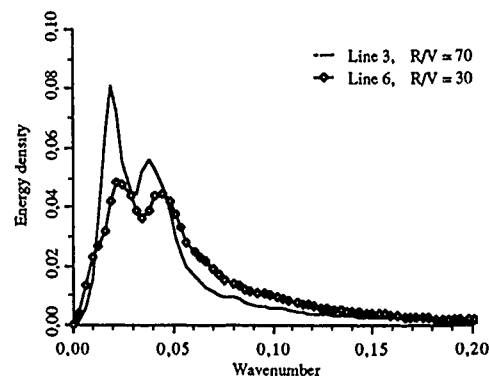


Fig. 7a. Heave spectrum obtained from Line 3 and Line 6 under different R/V ratios but same incidence angle $\theta=38^\circ$.

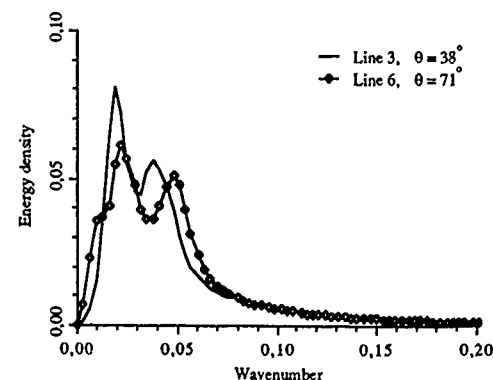


Fig. 7b. Heave spectrum obtained from Line 3 and Line 6 under different incidence angles but same $R/V=70$ s ratio.

Fig. 6 and 7 demonstrate the azimuth filtering as function of the R/V ratio. A change in R/V from 30 s to 70 s has a dramatic consequence on the directivity tending to move the peaks towards range. The heave spectrum is not that much influenced. The change in incidence angles has minor influence on the results. A quantitative measure of the azimuth filtering is obtained by

computing the azimuth profiles over the peaks in the spectra. Fig.8a shows the azimuth profiles at spectral peak mode 3 obtained from Line 3 and Line 6 under different R/V ratios but same incidence angle. The results of dividing the profiles in fig.8a according to eq. (6) is shown in fig. 8b. Fig. 9a and b shows the same plots obtained from modes 1 and 2.

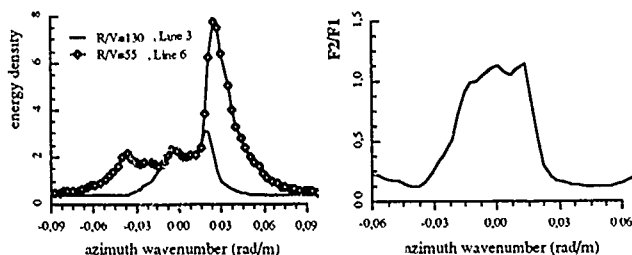


Fig 8a Azimuth profile at spectral peak (mode 3) obtained from Line 3 and Line 6 under different R/V ratios. Incidence angle $\theta=65^\circ$

b Results obtained by division of the profiles in fig.a, F_2/F_1 according to eq. (6) where F_2 is the profile obtained with R/V=130 s and F_1 with R/V=55 s

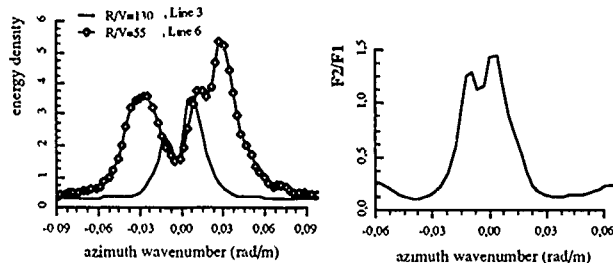


Fig 9a Azimuth profile at spectral peak (mode 1, 2) obtained from Line 3 and Line 6 under different R/V ratios. Incidence angle $\theta=65^\circ$.

b Results obtained by division of the profiles in fig.a, F_2/F_1 as proposed in eq. (6) where F_2 is the profile obtained with R/V=130 s and F_1 with R/V=55 s.

The half-width δk_a of the curves in fig.8b and fig.9b are estimated assuming a Gaussian shaped function according to eq. (6). The half-width δk_{a2} of the corresponding azimuth filter is then computed using the relation:

$$\delta k_{a2}^2 = \delta k_a^2 \left[1 - \left(\frac{R_1 V_2}{R_2 V_1} \right)^2 \right] \quad (7)$$

Obtained value for the half-width is $\delta k_{a2} \approx 0.013$ rad/m giving a minimum detectable azimuth wavelength $\lambda_{\min} = \frac{2\pi}{2\delta k_{a2}} \approx 240$ m for a R/V=130 s and a waveheight $H_s \approx 4$ m. This R/V ratio corresponds to the ERS-1 value. The λ_{\min} is in good agreement with theoretical predicted values (Monaldo, 1983). The coherence time can now be computed using eq. (3) and eq. (4), and obtained value is $\tau_{coh} \approx 0.015$ s. Fig. 10 shows the azimuth profiles over mode 1 and 2 obtained from Line 3 and Line 6 under same R/V ratio but under different angles of incidence.

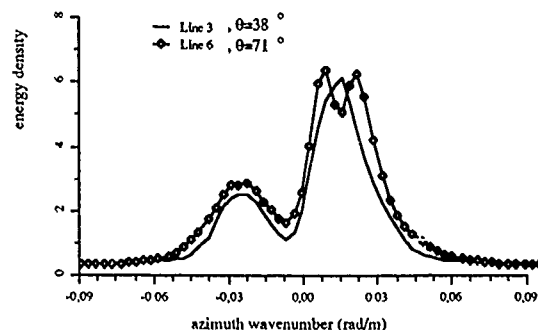


Fig.10 Azimuth profile at spectral peak obtained from Line 3 and Line 6 under different incidence angles but same R/V=70 s ratio.

The similarity between both profiles is very good both in shape and absolute energy density. The Sensitive Time Control (Livingstone, 1987) option was active during the generation of the real-time SAR data. This STC performs among several things an adjustment of the received power with respect to the variation of the radar reflectivity with incidence angle according to a radar reflectivity law. This combined with the careful preprocessing of the subimages as described in chap. 2 are probably the explanation of the results in fig.10. An increase in the angle of incidence leads to an increase in the integration time as specified in tab.2. Fig.10 indicates that the acceleration contribution being proportional to the integration time is negligible as a source to the observed azimuth filtering. Line 1, 2, 5 and 7 have also been investigated using the same procedure showing the same results regarding the azimuth filtering problem.

5. CONCLUSIONS

Loss of resolution in the azimuth direction is investigated as function of imaging geometry. The most important parameter is the range to velocity ratio, R/V, and a procedure for estimating an azimuth cut-off wavenumber and coherence time from SAR image spectra is proposed and demonstrated. For an R/V=130 s corresponding to the ERS-1 value, a cut-off in azimuth wavelength was estimated to 240 m. The coherence time was found to be of the order of 0.015 s. The azimuth filtering is shown to influence both the directivity and the peak wavelength of the ocean image spectra.

6. REFERENCES

1. Barstow, S.F., Bjerken, S., "Wave Measurement from Moored Directional Wave Buoys during NORCSEX '88", ODAP report no. 108, 1988.
2. Beal, R.C. et. al., "Large and Small-Scale Spatial Evolution of Digitally Processed Ocean Wave Spectra From SEASAT Synthetic Aperture Radar", J. of Geo. Res., Vol.88, No.C3, pp1761-1778, 1983
3. Hasselmann, K., et. al., "Theory of Synthetic Aperture Radar Ocean Imaging. A MARSEN View", J. of Geo. Res., Vol.90, No.C3, pp4659-4686, 1985
4. Livingstone, C.E., et. al., "CCRS C-band Airborne Radar - System Description and Test Results", Proceedings 11th Canadian Symp. on Rem. Sensing, Ontario, 1987.
5. Monaldo, F.M., Lyzenga, D.R., "On the Estimation of Wave Slope- and Height -Variance Spectra from SAR Imagery", IEEE Trans. on Geo. and Rem. Sensing, Vol.GE-24, No.4, pp543-551, 1983.
6. Tucker, M.J., "The Imaging of Waves by Satelliteborne Synthetic Aperture Radar. The Effects of Sea-Surface Motion", Int. J. of Rem. Sensing, Vol.6, No.7, pp1059-1074, 1985.
7. Vachon, P.W., et. al., "Airborne SAR Imagery of Ocean Surface Waves Obtained During LEWEX: Some Initial Results", IEEE Trans. on Geo. and Rem. Sensing, Vol.26, No.5, pp548-561, 1988.

AUTOMATED TRACKING OF ARCTIC ICE FLOES IN MULTITEMPORAL SAR IMAGERY

R. McConnell, W. Kober, F. Leberl
Vexcel Corporation
2477 55th Street
Boulder, CO 80301, USA
Tel: 303/444-0094
FAX: 303/444-0470

R. Kwok, J. Curlander
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

ABSTRACT

We discuss the use of two algorithms to perform shape matching on the boundaries of ice floes in SAR images in order to produce an ice motion map. The algorithms match a shape descriptor known as the psi-s curve. The first algorithm uses normalized correlation to match the psi-s curves, while the second uses dynamic programming to compute an elastic match that better accommodates deformation of the ice floe boundary.

I. Introduction

There has been considerable interest in recent years in the use of SAR imagery to study movement of arctic ice floes. The movement of ice floes is of interest to shipping, oil drilling, and military operations. In addition, the motion of ice floes plays a large role in the world's weather, because this motion exposes large expanses of unfrozen ocean water to the much more frigid arctic air, and is therefore responsible for a large amount of heat transfer between the ocean and the atmosphere. SAR imagery is suited for this task because it allows continuous coverage through clouds that prevail in the arctic, as well as during the dark winter months.

A common approach to the automated tracking of arctic ice has been to select a patch from an early image (the source image) and to cross correlate it with a later image (the target image) at each position that could plausibly correspond to the same patch of ice (Fily, 1986). The position that maximizes the computed correlation coefficient is deemed likely to contain the corresponding patch of ice. Sometimes consistency checks between several matches are used to identify false matches.

This method is known as area correlation. The problem with it is its computational expense, especially when ice floes rotate. To accommodate rotation, the patch must be rotated and correlated several times at each potential match position in the target image. The search space thus becomes very large, increasing the likelihood of false matches and increasing enormously the computational burden.

As a result, there has been increasing attention to algorithms for tracking ice floes by matching feature shapes (Vesecky, 1988).

We describe here a set of algorithms developed at jointly by Vexcel Corporation and the Jet Propulsion Laboratory (JPL) which have successfully overcome these obstacles. Rather than correlating raw pixel values, these algorithms first extract features from the images. The shapes of the extracted features are then compared using shape descriptors known as psi-s curves. To match the psi-s curves, the algorithms use both normalized correlation and dynamic programming. An ice tracking system that uses these algorithms is currently being developed jointly by Vexcel and JPL under a grant from NASA. The system, known as the Geophysical Processing System (GPS), is scheduled for installation at the Alaska SAR facility in Fairbanks in April of 1990.

II. Psi-s curves

Suppose $f(s) = (x(s), y(s))$ is a continuously differentiable parametric curve such that s is arc length along the curve. At each point, the vector $f'(s) = (x'(s), y'(s))$ gives the vector that is tangent to the curve.

Let $\theta(s)$ be equal to the angular difference between $f'(s)$ and the unit vector $(1, 0)$. θ is a real-valued, function of s which is continuous everywhere, except where phase wrapping introduces discontinuities, i.e., where it wraps around from 2π to 0 or from 0 to 2π . The psi-s curve is derived by adding or subtracting multiples of 2π to portions of the θ -s curve as needed to remove these discontinuities. In the case where the curve has points where it is not differentiable, multiples of 2π are added or subtracted in order to minimize the magnitudes of the discontinuities.

The boundary of a binary region in a digital image has a limited number of orientations, depending on the tessellation. It is therefore necessary to interpolate a smooth boundary through the jagged artifacts of digitization before computing the psi-s curve.

III. Matching of psi-s curves using correlation

One of the methods GPS ice tracking system uses for matching shapes is normalized correlation of the psi-s curves of the shapes. The GPS ice tracking system employs correlation by extracting arbitrary segments of fixed size from the psi-s curves from the source image and finding the subsegment of the same size in the psi-s curves from the target image that maximizes the correlation coefficient.

Correlating psi-s curves is an effective method of matching features that rotate strongly from one image to another. The reason is that rotation of a feature causes a constant to be added to its psi-s curve. The correlation coefficient between two functions is invariant to addition of a constant to one of the functions. Therefore, the ability of correlation to identify a match is unhampered by rotation of the features. The rotation can be estimated at the best match by linear regression of the psi values from the two sets.

Correlation is also invariant to scalar multiplication of one of the sets of samples being correlated. This occurs when one of the psi-s functions can be derived from the other by scalar multiplication with a factor other than 1.0. This would mean that the variation in the orientations of the tangents to one feature is greater than it is for the other, and hence, one of the features is a coiled up version of the other. This does not correspond to what humans would consider resemblance, nor are two such curves likely to represent the same ice floe in two images. The scalar multiple can be estimated the linear regression slope coefficient, and matches rejected where this estimate is significantly different from 1.0.

The final task is to determine which of the matches obtained by correlation are false and which are correct. The solution to this problem used in the GPS starts with the observation that most ice floes have at least two correct matches on them. The features involved in these matches move as part of a rigid body. The length of the axis joining them will be unchanged from one image to the next, and the estimates of the rotations of each obtained by their linear regression intercepts will coincide with each other and with the rotation of the axis joining the two features. It is unlikely that a pair of matches that are not both correct will satisfy these criteria. The GPS examines all pairs of matches obtained from psi-s correlation, tests for these criteria, and keeps all pairs of matches that satisfy them. This is a much more effective way of separating false matches from correct matches than is thresholding the correlation coefficients of the matches.

Figures 1, 2, 3, and 4 depict an ice image pair, the extracted features from each pair, and the motion vectors derived using psi-s correlation followed by the bad match filter.

IV. Problems with matching psi-s curves using correlation

Figure 5 illustrates a pairing between corresponding elements of two psi-s curves derived from segmented ice images. It is clear that the mapping of the elements is not one-to-one, and the mapping of the elements cannot be known until the curves are matched up.

In fact, implicit in correlation is the assumption that the psi values are disturbed from one image to the next, but that the mapping between the s domains remains linear. However, any disturbance of the psi values usually results from disturbance of the shape being matched, and therefore also results in distortion of arc length over some intervals of the boundary. The correlation matching process is therefore misspecified, and this can lead to failure of correlation matching to find correct matches when distortion of arc length is severe.

V. Dynamic Programming

The field of sequence comparison deals with the comparison of similar sequences, where the correspondence between the elements is not one-to-one and not known in advance. The techniques developed in this field provide a way of measuring the similarity between such sequences, as well as computing the optimal and most natural mapping between the elements of one sequence and those of the other. Most of the techniques in this field are based on a class of algorithms known as "dynamic programming". Dynamic programming is a process whereby a recursive problem with an exponential search tree can be solved in polynomial time by using a table to retain intermediate results that are shared by different branches of the tree. A survey of this area can be found in (Sankoff, 1983).

Dynamic programming has been used in the past for shape comparison in computer vision, but this has been restricted largely to handwriting analysis (Burr, 1983).

A. A Dynamic Programming Solution to the Matching of Psi-s Curves

A procedure known as "dynamic time warping" is a variant of dynamic programming for matching real-valued sequences. Dynamic time warping examines all sets of mappings between elements of one sequence and those of the other, subject to the constraints that the mappings do not cross each others, and that every element from one sequence is paired with at least one element from the other sequence. It assigns to each of these mappings the sum of absolute differences of paired elements. It then produces the mapping with a minimum sum of absolute differences, or "cost."

This can be computed recursively as follows:

Let x_n and y_n be two sequences of length m and n . Then x_i and y_i are the prefixes of these sequences containing i and j elements, respectively, and x_i and y_j are the i 'th and j 'th elements of the respective sequences.

The distance between the two sequences can be expressed as follows:

$$d(x_m, y_1) = \varepsilon |x_1 - y_1|$$

$$d(x_1, y_n) = \varepsilon |x_1 - y_n|$$

$$d(x_m, y_n) = \min \begin{pmatrix} d(x_{m-1}, y_n) - |x_m - y_n| \\ d(x_{m-1}, y_{n-1}) - |x_m - y_n| \\ d(x_m, y_{n-1}) - |x_m - y_n| \end{pmatrix}$$

The relation maps x_m to y_n at cost $|x_m - y_n|$, and resorts to recursion to compute the minimum-cost mappings of the preceeding elements.

The mapping can be computed cheaply by keeping a table, where element (i, j) of the table contains $d(x_i, y_j)$. By starting with element $(1, 1)$ the table can be built up inductively without resorting to recursion. When the table is completed, element (x_m, y_n) contains the cost of matching the two sequences. Backtracking in the array from element (x_m, y_n) through the elements giving rise to the minimum at each step of the recurrence relation gives the mapping between the elements. The time and space required to complete the table is proportional to its size, or $m \times n$.

The GPS uses a variant of this approach. The following recurrence relation is used:

$$d(x_m, y_1) = \varepsilon |x_1 - y_1|$$

$$d(x_1, y_n) = \varepsilon |x_1 - y_n|$$

$$d(x_m, y_n) = \min \begin{pmatrix} d(x_{m-1}, y_n) - r * |x_m - y_n| \\ d(x_{m-1}, y_{n-1}) - |x_m - y_n| \\ d(x_m, y_{n-1}) - r * |x_m - y_n| \end{pmatrix}$$

where r is a parameter that penalizes excessive arc-length warping.

After the table is completed, the bottom row is searched for a minimum value. Backtracking from this location gives the interval over x_m that matches y_n the best. When y_n is a fragment of a psi-s curve from one image, and x_m is the set of psi-s curves from the other, this procedure finds the best match.

Figure 6 illustrates the use of the table on a small example, while Figures 7 and 8 show dynamic programming match results on ice image data.

B. Problems with the Dynamic Time Warping Approach

As explained above, the primary advantage of the dynamic time warping approach over the correlation approach is that it is not blinded to a match by distortion of the s dimension of the psi-s curve. The chief disadvantage of the approach is that the underlying similarity measure is based on the sum of differences. Unlike correlation, the sum of differences measure of similarity is sensitive to the addition of a constant to

one of the sets of samples. In psi-s matching, this is exactly what happens when there is rotation of the features from one image to the next. Therefore, although the approach is not blinded by distortion in the s dimension of the psi-s curves, it can be blinded by strong rotation of the features to be matched.

To remedy this problem, a rotation-invariant descriptor of a curve can be derived from the psi-s curves by subtracting a running mean from them before the dynamic programming match is performed.

VI. Conclusion

Two methods of tracking arctic ice floes SAR images have been presented. Both of the methods do shape matching on psi-s curves. The first method, based on normalized correlation, is indifferent to rotation of the ice floes, and therefore excels in matching features whose rotation is unknown.

The second method, based on dynamic time warping, excels in matching features that resemble each other less closely, but whose rotation can be estimated.

REFERENCES

1. Burr, D.J., "Designing a Handwriting Reader", IEEE Trans. Patt. Anal. Mach. Intelligence, Vol. PAMI-5, No. 5, pp554-559, 1983.
2. Fily, M., and D. A. Rothrock, "Extracting Sea Ice Data from Satellite SAR Imagery", IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-24, No. 6, pp849-854, 1986.
3. Sankoff, D., and J. B. Kruskal, eds., Time Warps, and Macromolecules: the Theory and Practice of Sequence Comparison. Addison Wesley: 1983.
4. Vesocky, J.F., et. al., "Observation of Sea-Ice Dynamics Using Synthetic Aperture Radar Images: Automated Analysis", IEEE Transactions on Geoscience and Remote Sensing, Vol. 26, No. 1, pp38-47, 1987.



Figure 1. October 1978 Seasat SAR image of an area of central pack ice in the Beaufort Sea.

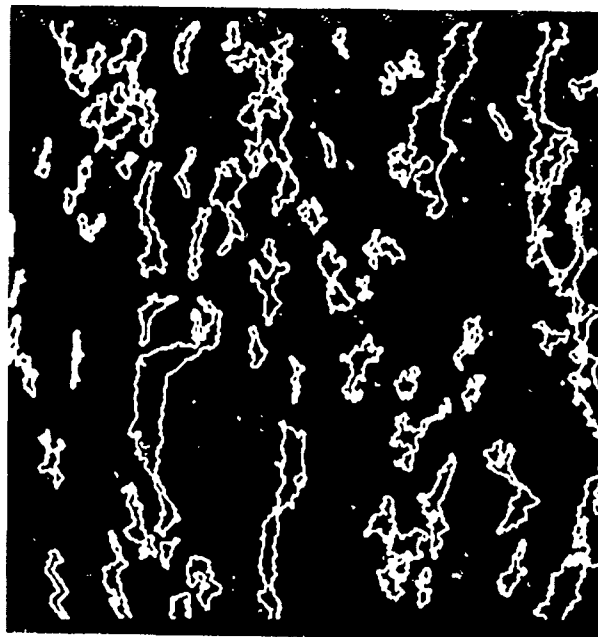


Figure 3. Features extracted automatically from the image in Figure 1. To extract the features, the gray values are clustered to get a binary classification. The boundaries of the regions in the resulting binary image are vectorized, and all boundaries with less than a minimum perimeter are discarded.

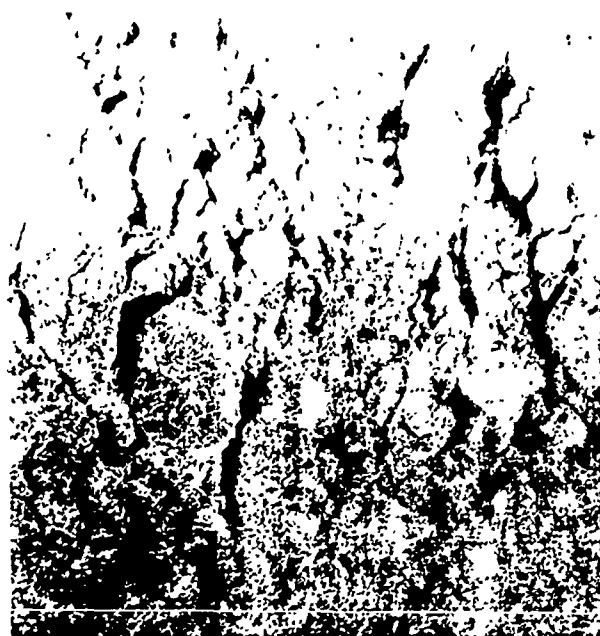


Figure 2. Seasat SAR image of the same ice three days later.

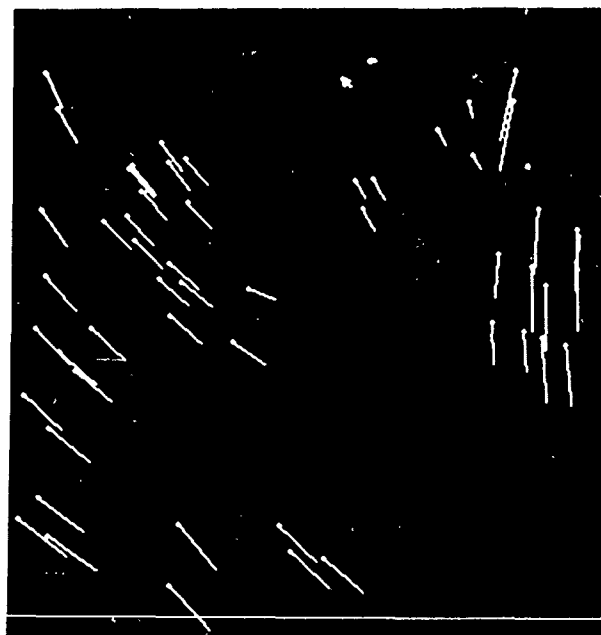


Figure 4. Correct matches obtained by correlation of the psi-s curves of features extracted from the images in Figures 1 and 2.

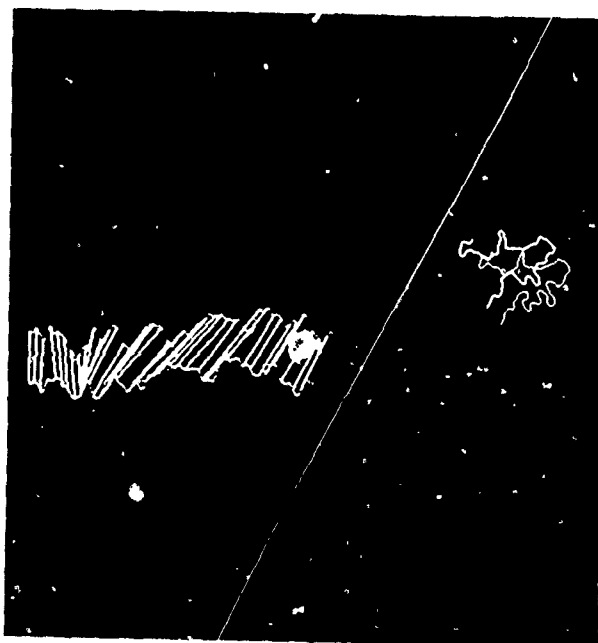


Figure 5. Two matching segments of sea ice feature boundaries and their psi-s curves. The matches of the psi values were produced using the GPS's dynamic programming algorithm. Deformation of the axis along which the curves match frequently inhibits a correct match when correlation of psi-s curves is used.

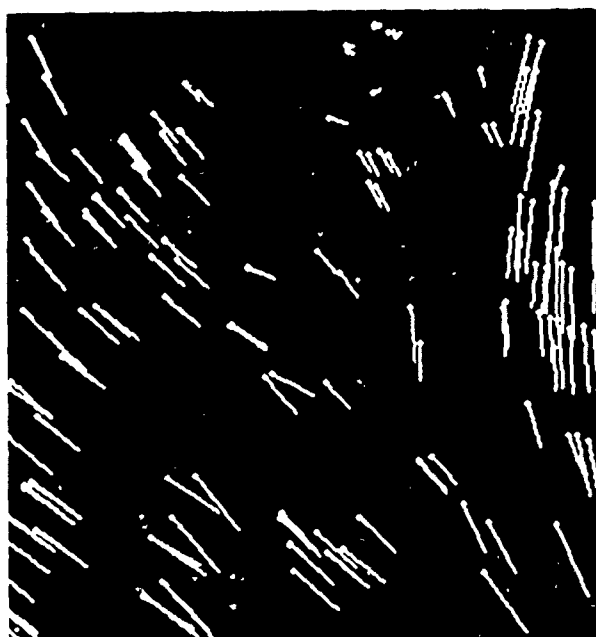


Figure 7. Matches produced by dynamic programming from the data used in Figure 4.

| | | | | | | | | | | | |
|---|----|---|---|---|---|---|---|---|---|---|----|
| | 1 | 2 | 1 | 2 | 3 | 3 | 2 | 5 | 0 | 1 | 1 |
| 2 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 3 | 2 | 1 | 1 |
| 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 | 2 | 5 | 3 | 3 |
| 2 | 4 | 1 | 2 | 1 | 1 | 1 | 0 | 3 | 4 | 4 | 4 |
| 4 | 7 | 3 | 4 | 3 | 2 | 2 | 2 | 1 | 5 | 7 | 7 |
| 4 | 10 | 5 | 6 | 6 | 3 | 3 | 4 | 2 | 5 | 8 | 10 |
| 0 | 11 | 7 | 6 | 8 | 6 | 6 | 5 | 7 | 2 | 3 | 4 |

Implied mapping:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 2 | 3 | 3 | 2 | 5 | 0 | 1 | 1 |
| 2 | 3 | 2 | 4 | 4 | 0 | | | | | |

Figure 6. Illustration of the table produced by a dynamic programming match of the sequence 232440 against the sequence 12123325011, which produces the best-match subsequence 233250. The table is filled in row by-row using the recurrence relation used by the GPS and a value of 1.0 for r . When the table is completed, the bottom row is scanned for a minimum value. Scanning vertically from the minimum value of 2, one finds the location of the end of the best-match subsequence. By backtracking from that location through the elements giving rise to the minimum in the recurrence relation, one finds the beginning of the best-match subsequence, as well as the mapping between the elements.

AUTOMATED ANALYSIS OF POLAR SATELLITE IMAGERY

Jeff Banfield*

Department of Mathematical Sciences
Montana State University

Drew Rothrock

Applied Physics Laboratory
University of Washington

Abstract

We have developed a technique that incorporates several innovative ideas to identify ice floes automatically. Our technique finds whole floes, not just partial edges or lead boundaries. Using standard image processing techniques, image pixels are classified as ice or water and edge pixels (those which define the border between ice and water) are identified. The ice floes are then eroded, using a computer to simulate melting the ice. The locations of those edge pixels which outline a given floe are propagated into the interior of the floe as it melts. This erosion-propagation (EP) algorithm produces initial clusters of edge pixels which outline the floes. It also eliminates loose ice and pixels which were erroneously classified as ice. The EP algorithm is computationally efficient and has the potential of being implemented on parallel processing machines. A new approach to cluster analysis, based upon principal curves and maximum likelihood estimation, is used for the final verification and representation of the floes.

Keywords: Ice Floes, Cluster Analysis, Erosion, Principal Curves

1 INTRODUCTION

With the advent of satellite imagery virtually unlimited observational data are available to study the movement and spatial structure of sea ice. In order to utilize fully this influx of data there is a need for automated pattern recognition techniques specifically designed for the analysis of sea ice. Many of the recent advances in the fields of

statistics, pattern recognition and computer vision are applicable to the analysis of remotely sensed sea ice. Using several of the new ideas from these fields, we have developed and implemented an algorithm which can take a satellite image and return the outlines and spatial configuration of all major ice floes. The shape and spatial information provided by our algorithm can be used in studying temporal and spatial variability in ice morphology, in floe tracking and in modeling ice processes such as lead dynamics and floe breakup. The solution to this complex problem requires a combination of several techniques, including cluster analysis and principal curve estimation. There are a number of preprocessing steps that must be undertaken before the clustering can take place. However, as can be seen in Figures 1 and 2, the results are quite satisfying.

In general, grayscale images contain more information than binary images. However, since the identification of ice floes is basically a binary problem it is reasonable to start by classifying pixels as ice or water. There are a number of contextual classifiers in the literature (Owen, 1984; Kittler & Illingworth, 1985), however the most common classification technique for creating binary polar images is thresholding (Vesecky et. al., 1988; Fily and Rothrock, 1986). In a thresholded image of sea ice, all pixels with an intensity above a specified threshold level are classified as ice, those below the threshold are classified as water. Although many pattern recognition techniques are sensitive to pixel mis-classifications due to thresholding, the one proposed here has proven robust to that problem. Setting the threshold level is one place where human supervision has proven fruitful. By allowing an operator to ad-

*Supported by ONR Grant N00014-89-J-1114

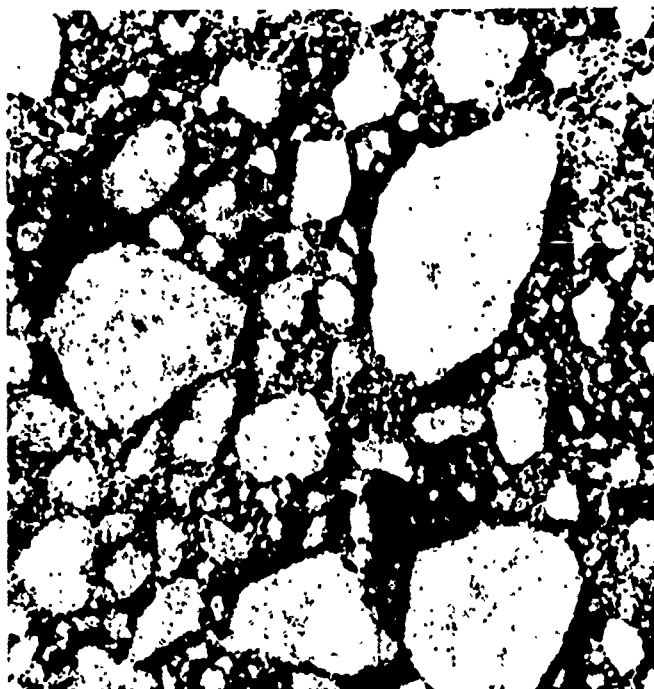


Figure 1. Ice floes in a LANDSAT image.

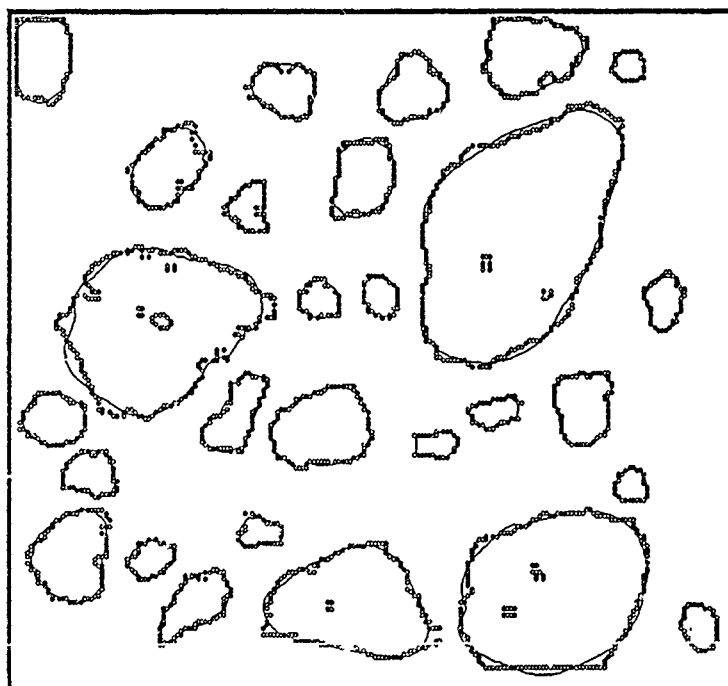


Figure 2. Ice floes that our method found in Figure 1. The circles are edge pixels, the smooth curves are principal curves that have been fit to each of the floes. The four largest floes have edge pixels in their interior. Because of the interior edge pixels these floes were subdivided by the EP algorithm. The partial floes resulting from the subdivision were then merged to form the complete floes in this figure.

just the threshold level interactively, the human vision system, with its exceptional pattern recognition capability, can very accurately determine an appropriate threshold level. If a fully automated system is desired, one of the more complex classification techniques can be employed.

After classifying the image as ice or water, the next step is to identify the edges of individual floes. This is done by determining which pixels are at the edge of a floe and then grouping together those edge pixels which belong to a common floe. Locating edge pixels is easy, there are many edge operators in the literature which can be used (Rosenfeld and Kak, 1982; Rosenfeld, 1984). Grouping together edge pixels which outline one particular floe is far more difficult. This is one of the problems that our algorithm addresses. Note that in Figure 2 the floe outlines are not just hundreds of individual edge pixels, they represent 28 objects, each one containing the outline of a single floe. All the edge pixels from a single floe are grouped together and know about each other (the ordering in which they fall, their distance from each other, their distance from the center, etc).

Our algorithm assumes that the edge pixels of a single floe must lie along a closed curve. By employing a clustering criterion based on closed principal curves (Banfield, 1988) it is possible to identify the resulting clusters as floe outlines. Unfortunately, there are generally so many edge pixels and so much noise that a direct application of cluster analysis is not practical. We produce a preliminary grouping of the edge elements by using erosion, a technique from mathematical morphology, as a thinning algorithm. Erosion simulates melting the ice and is able to locate a central point on each floe and associate that point with the appropriate edge pixels.

2 THE EP ALGORITHM

The basic idea behind erosion (Serra, 1982) is to define a structuring element and use it to remove pixels. To erode an ice floe the structuring element is placed over each pixel in the floe, if the structuring element lies entirely within the ice floe then that pixel "survives" to the next iteration. Otherwise, the pixel is eroded, that is, it is reclassified as water. A structuring element is simply a pattern that describes the neighbors

of a pixel. We used a 3x3 square, this defines the neighbors of a pixel to be the eight immediately adjacent pixels. By using other structuring elements different types of erosion and different erosion rates can be produced.

If a pixel is eroded and has no neighboring pixels that are classified as ice, then it must be the last pixel of a floe. By recording the location of the last ice pixel from each floe, an enumeration of the floes as well as the location of the "centers" of the floes is obtained. The minimum size of the floes can be controlled by waiting until a specified number of iterations have passed before recording the locations.

We define edge pixels to be the set of pixels removed by the first iteration of the erosion process. This set of pixels includes noise, the edges of melt ponds and loose ice as well as the pixels that outline the floes. If the locations of the edge pixels are propagated inward as the floes melt, then the last piece of ice from each floe should have associated with it all the edge pixels from that floe. The edge pixels not associated with a floe will be eliminated. The iterative procedure of eroding a closed figure while propagating the edge information to the center is called the erosion propagation (EP) algorithm (Banfield and Raftery, 1989a). The EP algorithm produces a collection of objects that may be floes. Unfortunately, the EP algorithm can subdivide floes that are non-convex or have dark areas (such as melt ponds) in their interior.

If a floe is subdivided, the resulting partial floes will have common edge elements. Unfortunately, separate floes which lie close to or touch each other can also share edge elements, so common edge elements alone cannot be used to determine which collections of pixels from the EP algorithm should be merged. It can, however, indicate which of the partial floes should be considered as candidates for a merger. In order to determine whether a set of floes which share edge elements should be merged we use a statistical technique known as cluster analysis.

3 CLUSTER ANALYSIS

The first step in determining whether adjacent floes should be combined is to fit a closed principal curve (Hastie, 1984; Banfield, 1988) to each of the collections of edge pixels produced by the

EP algorithm. The characteristic that allows the successful merger of the partial floes is that the edge elements of a complete floe have uniformly distributed projections onto the principal curve, while large gaps in the projections indicate a partial floe. We cluster the edge pixels about the principal curves using a clustering criterion based upon both the variability of the points about the curve and the variability of the projections along the curve. The criterion is

$$V = \alpha V_{\text{about}} + V_{\text{along}} \quad (1)$$

where V_{about} is the variance of the lengths of the projections of the data onto the principal curve and V_{along} is the variance of the distance, along the principal curve, between adjacent projections.

To determine if a set of adjacent floes should be merged we first calculate V for each of the individual floes, then V for the floe that would result from the merger of the edge elements of the individual floes. If the floe resulting from the merger has a smaller value of V than any of the individual floes, the merger is needed. Otherwise, the individual floes should not be merged.

In order to calculate V an estimate of α is needed. It is shown (Banfield and Raftery 1989b) that a criterion of the form given in equation (1) is optimal (it is the maximum likelihood estimator) if, over all floes, $V_{\text{along}} = \alpha V_{\text{about}}$. Thus, a reasonable estimate of α is the average of

$$\frac{V_{\text{along}}}{V_{\text{about}}}$$

for all the floes which were not subdivided. Using the complete floes found by the EP algorithm, those floes which do not have any shared edge elements, we estimated α for the floes in Figure 1 to be .39

4 CONCLUSIONS

Figure 2 shows the principal curves and edge elements of the floes in Figure 1 after partial floes have been merged. The principal curves have been fit using a scatterplot smoother. They present a smooth outline of the floe shapes. Since they are smoothed versions of the edge elements, they tend to round sharp edges and corners. It is possible to get a more detailed outline by adjusting the scatterplot smoother. If the exact outline is needed the edge elements themselves can be

used. The edge elements in the central regions of the 4 largest floes result from melt ponds and pixel mis-classifications. They are not part of the floe edges and have been given zero weight in calculating the principal curves. They may easily be eliminated when working with the edge elements.

References

- [1] Banfield, J., "Constrained Cluster Analysis and Image Understanding," Ph.D. Dissertation, Statistics Dept, University of Washington, 1988.
- [2] Banfield, J. and Raftery, A., "Ice Floe Identification in Satellite Imagery" In preparation, 1989a.
- [3] Banfield, J. and Raftery, A., "Constrained Cluster Analysis," In preparation, 1989b.
- [4] Fily, M. and Rothrock, D. A., "Extracting Sea Ice Data from Satellite SAR Imagery," *IEEE Trans. Geosci. and Remote Sensing*, GE-24, 849-8540, 1986.
- [5] Hastie, T., "Principal Curves and Surfaces," Ph.D. Dissertation, Statistics Dept, Stanford University, 1984.
- [6] Kittler, J. and Illingworth, J., "Threshold Selection Based on a Simple Image Statistic," *Computer Vision, Graphics, and Image Processing*, 30, 125-147, 1985.
- [7] Owen, A., "A Neighbourhood-based Classifier for LANDSAT Data," *The Canadian Journal of Statistics*, 12, 191-200, 1984.
- [8] Rosenfeld, A., "Picture Processing: 1984," *Computer Vision, Graphics, and Image Processing*, 30, 185-242, 1985.
- [9] Rosenfeld, A. and Kak, A., *Digital Picture Processing, 2nd Ed.*, Academic Press, 1982.
- [10] Serra, J., *Image Analysis and Mathematical Morphology*, Academic Press, 1982.
- [11] Vesecky, J. F., Samadani, R., Smith, M. P., Daida, J. M. and Bracewell, R. N., "Observation of Sea-Ice Dynamics Using Synthetic Aperture Radar Images: Automated Analysis," *IEEE Trans Geosc. and Rem. Sens.*, 26, 38-48, 1988.

Object-Based Feature-Tracking Algorithms for SAR Images of the Marginal Ice Zone

Jason Daida & John Vesecky

STAR Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305-4055

Abstract

Synthetic Aperture Radar (SAR) provides an excellent means of observing the movement and distortion of sea ice over large temporal and spatial scales. Consequently, the European Space Agency's ERS-1 satellite will carry a SAR over the polar regions in late 1990. A key component in using arctic SAR data is an automated scheme for extracting sea-ice displacement fields from a sequence of SAR images of the same geographical region.

Although automatic sea-ice tracking algorithms do exist, analyzing the Marginal Ice Zone remains challenging. The wide variety of ice movements and diverse of seascapes have led to the development of hybrid schemes. An important element of these schemes consists of a feature-based algorithm. Another important element usually consists of a statistically-based algorithm, e.g. correlation. Computer understanding of what features to look for, when to apply correlation and where to look is still subject to investigation.

Our research focuses on a method for providing computer understanding of a Marginal Ice Zone scene. The method parses a complex ice scene into individual objects. We define an object as a closed boundary and its interior. Consequently, we highlight in this paper how the recognition of objects facilitates sea-ice tracking. We also describe our solution to the problem of weakly connected regions as one facet in automatically creating objects from bitmaps. We provide an example of matching floes taken from a synoptic image pair. The example employs a simple object construct using invariant moments, although object constructs are not limited to them.

1. Keywords

Marginal Ice Zone, Image Understanding, Object Identification, invariant moments, automated sea-ice tracking algorithms.

2. Introduction

Conditions found in the Marginal Ice Zone dictate that a tracking scheme account for a wide variety of ice movement and a diversity of seascape. Ice found close to pack ice will often move as rigid plates. Ice found close to open ocean will often move as fragments, especially during the summer. These fragments are difficult to track. Unlike rigid plates, fragments can spin many times between SAR observations as they drift across open water. The context within which fragments occur, e.g. an open-water/new-ice matrix, may change rapidly. Edges of fragments can change as a result of melt and collision. Fragments may also further disintegrate into smaller fragments. Figure 1 illustrates how some of these phenomena appear in an SAR image. Given the variety of conditions within the Marginal Ice Zone, multiple techniques should and can be employed to track ice.

Recent hybrid schemes have been introduced to accommodate a variety of ice movement. Vesecky et al. proposed a hybrid of feature / statistical-based tracking algorithms. (Vesecky et al., 1988) The Alaskan SAR Facility Geophysical Processor System implements a feature / statistical-based tracking scheme using the edges of segmented regions for features and pyramid correlation for statistical matching (Kwok, 1988). Generally, boundary features work with most fragments, while pyramid statistical correlations work with rigid plates.

The class of algorithms under investigation uses objects, which are amenable to feature tracking as well as statistical tracking. Object characteristics may indicate the type of algorithm to use; for example, large objects use pyramid correlation and small objects use features. *Objects may take into account interior features as well as their boundaries.* Matching objects representing fragments directly facilitates tracking. Finally, object understanding will help in integrating feature-tracking and statistical-tracking methods.

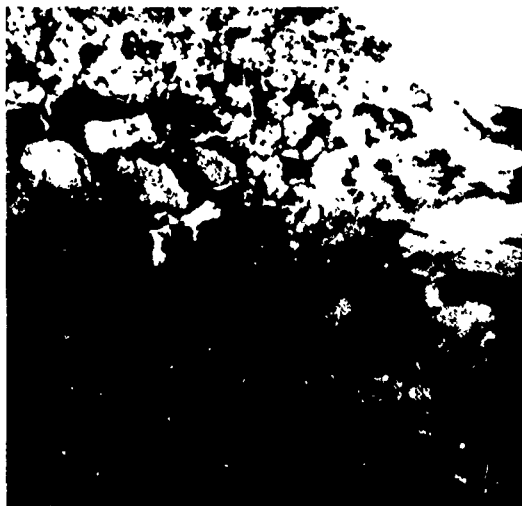


Figure 1. Marginal Ice Zone from Barrow, Alaska. This synoptic image pair was taken by SEASAT in 1978. Each pixel represents a 200 x 200 meter area. a) Orbit 1409 October 3. b) Orbit 1452 October 6.

3. Framework: Object Understanding

Objects are meta-features, collections of individual features belonging to a particular spatial group. We define an object as a closed boundary and its interior. Whereas edge-based algorithms use only the information given by the boundary, object-based algorithms use the information given by the boundary as well as the information given by its interior. An object may simply consist of a binary image with ones representing the region of interest and zeroes elsewhere. An object may also consist of a gray-scale image with intensity values representing the region of interest and zeroes elsewhere. An object may further consist of a feature-reduced image with the region of interest presented by a closed boundary and some interior features, such as pressure ridges. Whatever the variation, object-based tracking algorithms can use the information within the interior of a boundary to characterize a floe or region. The advantage comes when individual features, such as edges, are insufficient for registration and when the spatial context of those features becomes important.

A pseudo-code implementation in C of an object would look something like the following:

```
struct object
{
    list of coordinates of this object;
    an identification tag;
    number of pixels in this object;
    /* for sorting by mass */
    feature x
    /* e.g. ice type */;
    feature y
    /* boundary vector */;
    feature z
    /* e.g. pressure ridge density */;
    .
    .
    .
};
```

A flowchart of a tracking algorithm using objects could look like that shown in Figure 2.

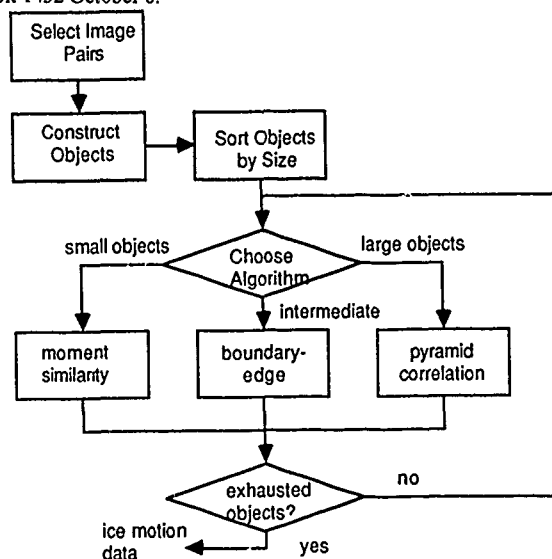


Figure 2. Example flowchart incorporating objects.

In this example, objects mediate the first level of decision making. Subsequent decisions based on objects can be made at a lower level to aid in determining what other algorithm to use.

4. Object Construction and Algorithm

The process of automatically constructing two-dimensional objects from a bitmap includes two general steps. The first step is to segment an image. These segments correspond roughly to ice (e.g., first, second, and multi-year types) and water (e.g., new ice, frazil, open water). Classification and segmentation techniques are fairly well discussed in literature. The next step is to identify contiguous regions which roughly correspond to floes. The latter step of identifying contiguous regions

requires further elaboration.

The problem of how to identify specific regions is simple conceptually. Each spatial cluster of pixels corresponds to a region. Each region is issued an identification tag. The problem would be simple if regions were non-connecting. However, in even highly fragmented scenes of ice, regions may be weakly connected.

To address the problem of weakly connected regions, we have used the following algorithm. Start with a binary image: ones for ice, zeroes for water. Create a list of coordinates corresponding to those pixels that have a value of one. Select a pixel from that list and label it active. Set that pixel's eight-connected neighbors as active. After determining the initial pixel's neighbors, designate it solid.

The next sequence is iterative. Using active pixels, determine the next layer of potential pixels. A potential pixel is eight-connected to an active pixel. A potential pixel is also on the ones-list and is not solid. Eliminate weakly connected potential pixels. A potential pixel is eliminated if and only if it fails to be connected to a certain number and orientation of solid or active pixels. Set the current active pixels to solid. Set the remaining potential pixels to active. Repeat and continue until no more potential pixels can be found. Figure 3 exemplifies this sequence.

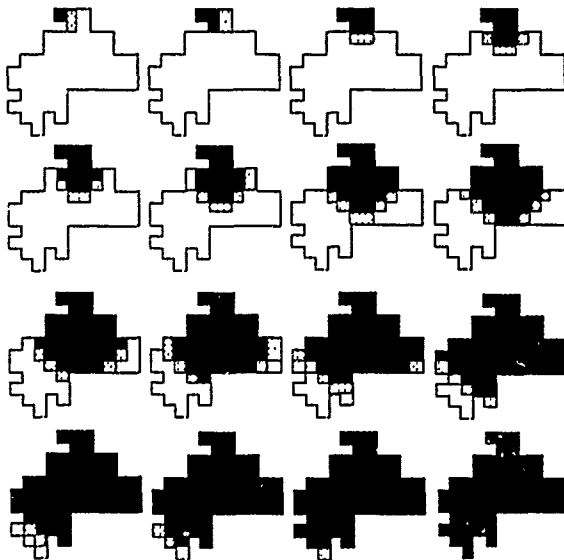


Figure 3. Identification process of a 62 pixel object. Solid pixels are determined as part of the object. Gray pixels are new locations that have been determined as part of an object. From the gray pixel location, new candidates for solid pixels are chosen.

After iteration, all pixels belong to a region. These pixels can now be tagged. Pixels that are tagged are removed from the ones-list. A new region is started from a location in the ones-list. The entire procedure repeats until no more pixel locations exist in the list. See Figure 4 for an example.

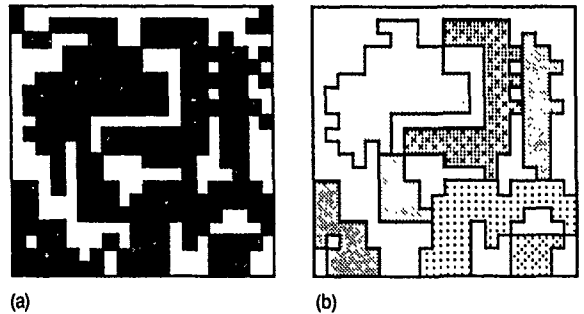


Figure 4. A 20 x 20 pixel example demonstrating weakly connected regions. (a) Binary image showing seven weakly connected objects. (b) Result of algorithm applied to (a). Seven regions are identified and each region is tagged with a particular intensity value. Note that objects of one-pixel width are filtered out completely. The object described in Figure 3 is part of (a).

The amount of inter-region filtering depends on what criteria are used to eliminate potential pixels from further consideration. The current algorithm requires that a potential pixel have two adjacent eight-connected neighbors that are either solid or active. Consequently, the algorithm defines weakly connected regions as regions connected by corner or by one-pixel bridges. More stringent criteria can be applied for greater inter-region filtering.

5. Application: Registration with Moments

A. Object Registration Algorithm with Invariant Moments

We selected a pair of 20 x 20 km subimages from SEASAT image pairs 1409 and 1452 for demonstration. The images were sub-sampled to create 100 m pixels to simulate ERS-1 data. Subsequently, quadtree-reductions smoothed the image to 200 m pixels. Subimages were then extracted. Binary images were created from the subimages with a simple segmentation based on thresholding a mean image generated by a 3 x 3 pixel window. Figure 5 shows the resultant binary image pair.

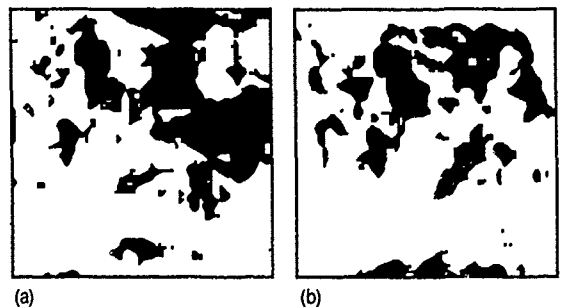


Figure 5. Binary subimages. (a) 1409 subimage of pair a. (b) 1452 subimage of pair b.

The identification algorithm was applied to both image pairs. The identifier distinguished 14 objects from Figure 5(a) and 19 objects from Figure 5(b) greater than 8 pixels in size. (See Figure 6.)

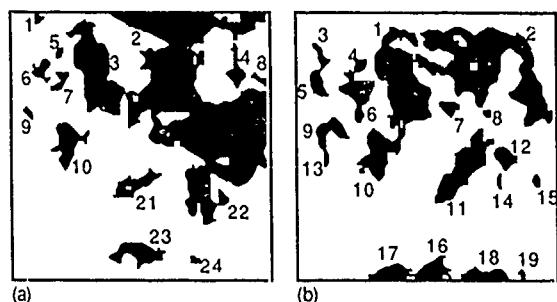


Figure 6. Identified Objects. Most of the salient features were retained in the process. (a) 1409 subimage. (b) 1452 subimage.

After identifying the location of each object, a tracking scheme could assemble and append other attributes. For example, we constructed the following object.

```
struct object
{
    list of coordinates of this object;
    an identification tag;
    number of pixels in this object;
    /* for sorting by mass */
    vector of invariant moments;
};
```

The vector of invariant moments is based on Hu's set of seven invariant measures (Hu, 1962). These measures are approximately invariant for digital images under translation, rotation, and scaling. A fast means of computing these moments on general-purpose computers is described by Zakaria, et al. (Zakaria, 1988).

Further filtering of objects was applied.

Ambiguous objects, usually less than 100 pixels in size, were removed from further consideration. Objects which intersected the image border were also removed.

The remaining objects were sorted by size. The size of objects remaining in the first image dictated the scope of search in the second image. Objects in the second image were roughly the same size as objects in the first image. In this example, we chose object sizes in the second image to be ± 50 percent of an object size in the first image.

Finally, the invariant moments of each object were compared. The similarity measure we used was

$$M = \sum_{i=1}^7 ||\log(\phi_1(i)) - \log(\phi_2(i))||$$

where ϕ is an invariant moment from either object 1 or 2.

B. Results

Two objects can be manually registered based on the binary image. These matching floes correspond to: Object 3 in 1409 and Object 1 in 1452; Object 10 in 1409 and Object 10 in 1452. See Figure 7.

The algorithm successfully registers both pairs. Using object mass for filtering, only one possible match is made between Object 3 (1409) and Object 1 (1452). The match is verified with a similarity measure of .83 (a perfect match is 0, wrong matches may span several orders of magnitude). Mass and border criteria narrow

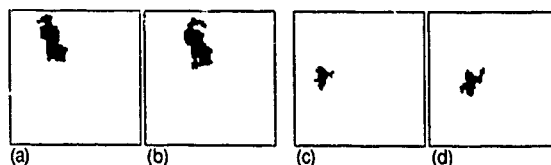


Figure 7. Matching Floes. (a) Object 3: 1409. (b) Object 1: 1452. (c) Object 10: 1409. (d) Object 10: 1452.

the list of possible matches to Object 10 (1409) to 3 choices. The similarity measure determines that Object 10 (1452) is a correct match.

6. Discussion

Ambiguities and missed hits will occur with the smaller objects. A limitation of how small an object can be matched is dependent on the performance of the classifier which precedes identification. For example, Objects 11 in both 1409 and 1452 are matched floes; however, the binary image of Object 11 in 1409 is very different from the binary image in 1452. Likewise, Objects 16-19 in 1452 are actually contiguous in the grayscale image. We would expect incomplete representations from the simple classifier used. A more rigorous classifier would help in keeping the ambiguities down.

The above example demonstrates a simple application of objects in problems of automatic sea-ice tracking in the Marginal Ice Zone. Objects were used to parse more complex scenes into parts which can be analyzed individually. Objects were also used to keep related features of mass and invariant moments organized and in context of one another. Finally, objects were used to match floes from one frame to another—a step which precedes creating a displacement map for tracking.

7. Acknowledgements

We are grateful to F. Carsey, J. Curlander, B. Holt (Jet Propulsion Laboratories) and R. McConnell (Vexcel Corporation) for providing the images; R. Samadani and A. Bernardi (Stanford STAR labs) for their expertise; Apprise, Inc. for making possible a Macintosh IIx; S. Obayashi for her support and helpful comments. This work was supported by NASA Oceanic Processes Branch (Robert Thomas) & ONR Remote Sensing (Charles Luther).

8. References

- [1] J. Vesecky, R. Samadani, M. Smith, J. Daida, & R. Bracewell, "Observation of Sea-Ice Dynamics Using Synthetic Aperture Radar Images: Automated Analysis," *IEEE Transactions on Geoscience & Remote Sensing*, vol. GE-26, no. 1, pp.38 - 47, January 1988.
- [2] R. Kwok, "GPS Review: Ice Motion Tracker: Status/Results/Performance," ASF SWT Plumbers Workshop, December 20, 1988.
- [3] M.K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inform. Th.*, IT-8, pp. 179-87, February 1962.
- [4] M. Zakaria, L.J. Vroomen, P. Zsombor-Murray, & J. van Kessel, "A Fast Algorithm for Moment Invariants," *Image Analysis and Processing II*, Plenum Press, New York, pp. 261-268, 1988.

A HOUGH TRANSFORM TECHNIQUE FOR EXTRACTING LEAD FEATURES FROM SEA ICE IMAGERY

Florence M. Fetterer and Ronald J. Holyer

Remote Sensing Branch
Naval Ocean Research and Development Activity
Stennis Space Center, MS, USA 39529

ABSTRACT

Infrared imagery from polar-orbiting satellites provides a synoptic and long-term view of lead patterns in Arctic pack ice. The large quantity of satellite data in image form suggests the use of automated methods for compiling lead statistics from imagery. A Hough transform technique for the semi-automated extraction of lead orientation and spacing is described and first results are presented.

KEY WORDS: SEA ICE, LEADS, HOUGH TRANSFORM, AVHRR

1. INTRODUCTION

Sensors on polar-orbiting satellites produce digital imagery with which the regional, seasonal, and annual variability of large-scale lead patterns in Arctic pack ice can be studied. The infrared AVHRR sensor on board the NOAA series is well suited to this task because of its nearly arctic-wide swath and revisit frequency of several times per day. Thermal data offers day and night, year-round coverage. However, cloud contamination is often a problem at high latitudes, and because AVHRR imagery has a resolution of 1 km at nadir, only large leads are resolved. Within these limitations, characteristics of a lead ensemble such as average lead width, length, spacing and orientation can be obtained from imagery and related to atmospheric and oceanic stress fields and to surface fluxes.

Satellite remote sensing provides the spatial and temporal coverage necessary to compile a climatology of lead characteristics. With this advantage comes the task of extracting lead features from large amounts of image data. Here a method for the semi-automatic extraction of lead orientation and spacing from binary images of leads is presented. Although demonstrated with AVHRR imagery the technique is suitable for any digital imagery in which lead pixels can be distinguished from background pixels.

2. THE HOUGH TRANSFORM

The Hough transform is a computer vision technique for detecting lines, circles, or other shapes in imagery. Recently it has found application in remote sensing (Cross, 1988), where its speed, flexibility in extracting shapes, and relative insensitivity to image noise makes it an attractive alternative to other methods for object recognition. To implement the transform, an edge detector or other method for identifying leads is first applied to the image.

Each pixel of the image is classified as either a lead pixel or not a lead pixel. Pixels identified as lead pixels in image space are then mapped into "parameter space", where parameters are those which describe the shape being sought. Here, it is assumed that lead pixels are roughly colinear points, therefore the transform is employed as a line finder using the normal parameterization of a line

$$\rho = x \cos \theta + y \sin \theta$$

where θ is the angle of a normal to the line and ρ is its distance from the origin (Fig. 1a). Parameter space is represented as an accumulator array of discrete θ, ρ values. For an (x, y) image pixel, θ is incremented from 0 to 180 in steps of one degree and ρ calculated for each θ . Accumulator element (θ, ρ) is incremented for each pair. Every lead point in the image is therefore transformed into a sinusoidal curve of 180 points in the accumulator, where each (θ, ρ) element of the curve describes possible lines through that image point. If points in image space are members of the same line, curves formed by those points will cross in parameter space at the (θ, ρ) element which describes the line (Fig. 1b). That element will have a higher value than surrounding elements in the accumulator array (Duda and Hart, 1972). The accumulator array can be thought of as a 2-d histogram of the frequency with which points occur on a given line. Peaks in the accumulator above a noise threshold correspond to lines in the image.

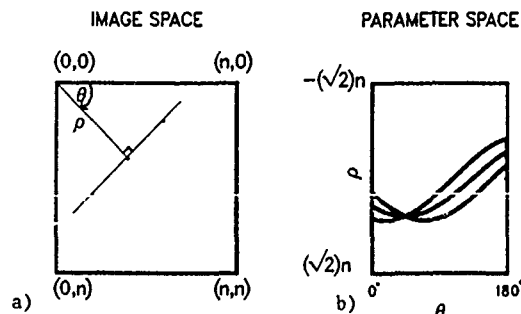


Fig. 1. The Hough transform maps points from image space (1a) into parameter space (1b). See text for explanation.

3. APPLICATION TO LEAD FEATURES

3.1 Test Images

Figure 2 shows the AVHRR image of the central Arctic from which test images 1-4 were selected. Image resolution varies with distance from nadir. For simplicity, we will assume a constant 1km per pixel. Leads are warmer than the surrounding ice, and therefore appear as dark lines in infrared imagery. Binary images were created from images 1-4 using an interactive detrending and thresholding procedure (Fig. 3b, 5a). This preprocessing step is critical, as the Hough transform will detect only those lines which appear in the binary image. Future work will seek to make this step less subjective and to preserve the distinction between new, dark leads and old, light leads which have a thicker ice cover. Fortunately, the Hough transform is insensitive to breaks in lead lines which are created in the thresholding process.

3.2 Transform Procedure

Fig. 4a displays the accumulator for the image of Fig. 3b in image form. The resolution of each pixel is one degree in the θ direction, and 2km in the ρ direction. Peaks, or bright points, in the accumulator give the angular orientation (θ) and normal distance to image origin (ρ) for leads in the image. Because lead pixels are only approximately colinear, peaks will be spread over a cluster of points. An empirically chosen threshold was applied to the accumulator, and points with values above the threshold (shown in Fig. 4b) were inverse-transformed to determine how well leads in the binary image were detected (Fig. 3c). Parameter space carries no information on the position of line end points, therefore calculated lines are drawn with infinite length.

The inverse transform of all points in clusters with values above the threshold results in a fan of lines for each lead in image space. A central point for each cluster in Fig. 4b was inverse-transformed to give an average representation of each lead (Fig. 3d). All prominent leads were detected, but several short or less prominent leads were not, because they do not have colinear points greater in number than the empirically chosen accumulator threshold of 51. The near-vertical line on the right in Fig. 3d is a false detection - at least 52 points lie on that line in image space. While lowering the threshold increases the probability of detecting all leads, the number of false detections rises also. Using smaller images mitigates this problem to some degree by lessening the chance that the number of randomly colinear points will exceed the number of points in the shortest lead. However, detection will always be biased toward longer (and straighter) leads.

In Fig. 5, inverse transforms for images 2-4 using accumulator cluster centers are shown. Thresholds were lowered until the first occurrence of a false detection, and then raised to prevent false detection.

4. RESULTS

In order to display parameter space information as lead orientation and spacing, histograms of the frequency with which points occur on an image line are plotted (Fig. 6). The orientation histogram is found by summing the elements in columns of an accumulator, such as that of Fig. 4b, in which points below the threshold have been set to 0. The normal length, or distance histogram, is found by summing

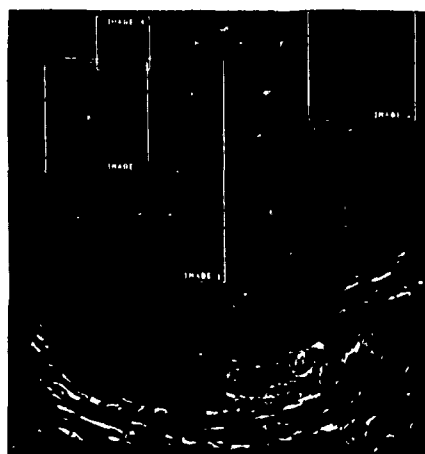


Fig. 2. NOAA-9 AVHRR image from 25 April, 1980, enhanced to highlight leads.

row elements. The distance histograms have peaks which are higher and sharper than those for orientation, in part because orientation may change greatly along a lead while the normal distance to the origin stays roughly the same. Both lead length and the number of leads having the same orientation or normal distance from the origin contribute to the histograms in Fig. 6. A smoothed version of the orientation histogram is plotted as a representation of average orientation. It is of note that for images 1 and 2, orientation peaks are separated by about 30 degrees. This is near the 28 degree intersection angle Marko and Thompson (1977) found characteristic of Arctic leads.

The spacing between peaks in the distance histogram is an approximate measure of lead spacing. The histogram is treated as a spatial series and a Fourier transform performed. Peaks in the smoothed amplitude spectrum plotted to the right of each distance histogram correspond to lead spacings. Unlike taking an average of lead spacings, this method distills spacing information while preserving the bimodal nature of lead spacings such as those in image 1.

5. CONCLUSIONS

The Hough transform simplifies the analysis of images for lead statistics. It yields lead orientation and spacing information quickly and in a form which is easy to interpret. As demonstrated here, it is an incomplete method in that some pre-processing is required to produce a binary image, and post-processing must be done to extract average lead width and length. Within these limitations, the Hough transform permits semi-automated extraction of lead parameters from digital imagery. Future work will attempt to fully automate the procedure by setting the accumulator threshold using criteria to avoid false detections (e.g. Gerig, (1987)). In addition, the statistical accuracy of the method must be established.



Fig. 3. Original image 1 (3a), binary image 1 (3b), inverse transform of accumulator points above a threshold of 51 (3c), and inverse transform of accumulator cluster centers only (3d).

6. REFERENCES

- Cross, A.M., "Detection of circular geological features using the Hough transform", *Int. J. Remote Sensing*, Vol. 9, No. 9, pp1519-1528, 1988.
- Duda, R.O. and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures", *Commun. ACH*, Vol. 15, No. 1, pp11-15, 1972.
- Gerig, G., "Limiting image-space and accumulator space: a new approach for object recognition", in *Proc. 1st Int. Conf. Comput. Vision*, pp112-117, June 8-11, 1987.
- Marko, J.R., and R.E. Thomson, "Rectilinear leads and internal motions in the ice pack of the western Arctic ocean", *J. Geophys. Res.*, Vol. 82, No. 6, pp979-987, 1977.

Approved for public release; distribution is unlimited. NORDA contribution No. PR 89:034:321.

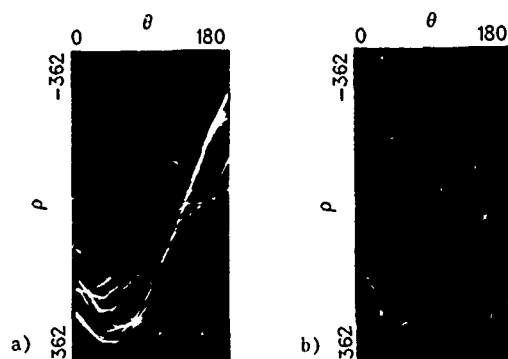


Fig. 4. Image 1 accumulator (4a) and accumulator with threshold applied (4b).

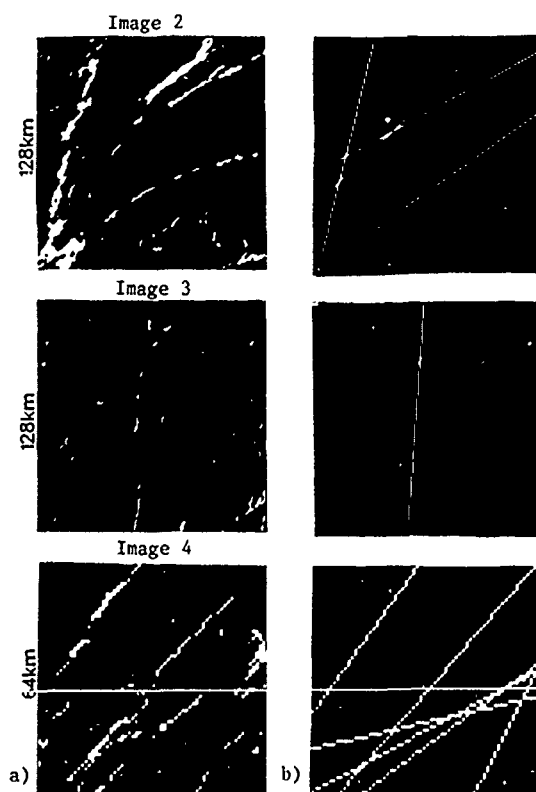


Fig. 5. Binary lead images (5a) and detected leads from accumulator peaks (5b) for images 2-4.

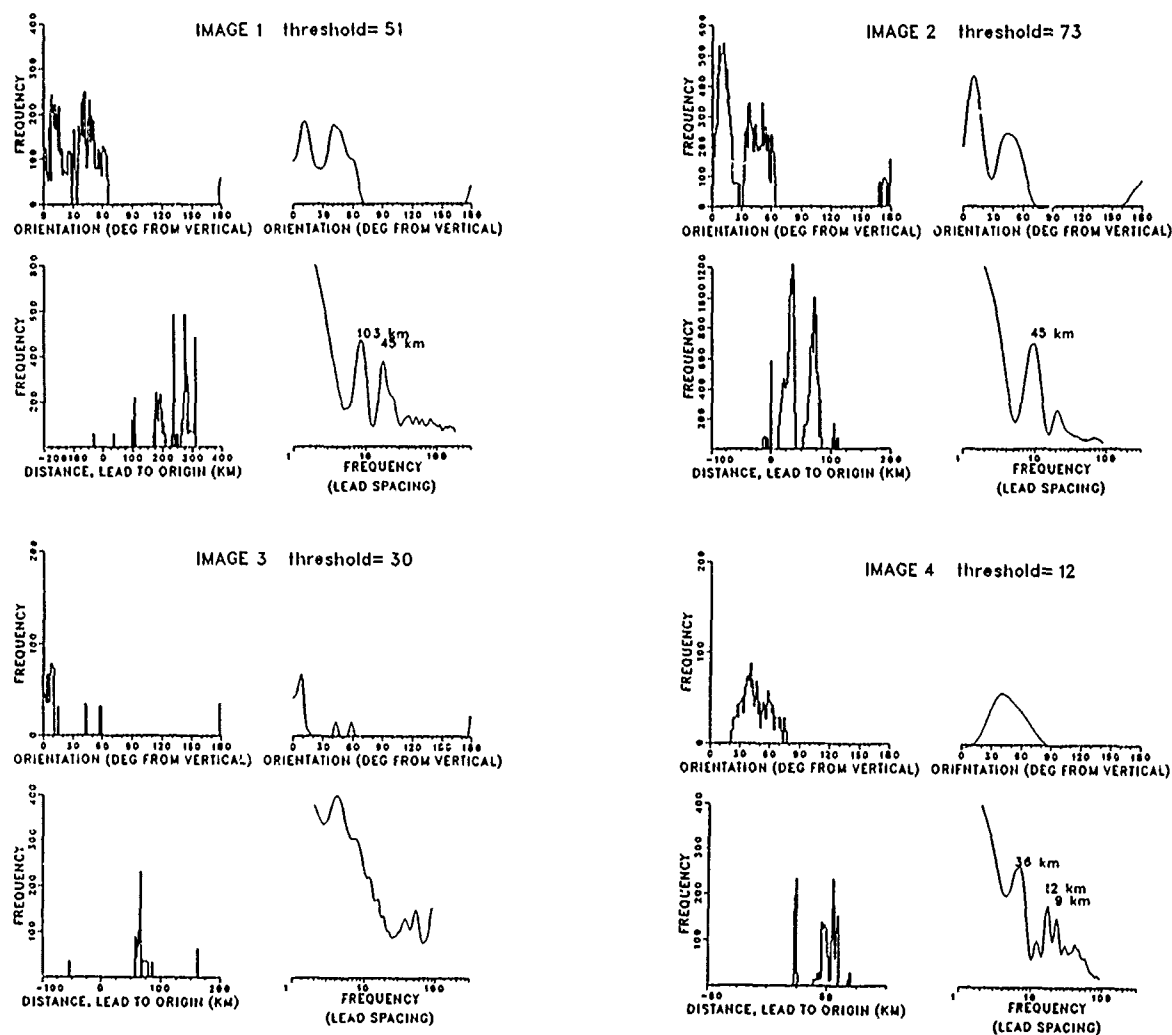


Figure 6. Lead orientation and spacing from Hough transform parameter space. Histograms on left, smoothed orientation histograms and smoothed Fourier transform of distance histograms on right.

EXTRACTION OF RIDGE FEATURE CHARACTERISTICS FROM SAR IMAGES OF SEA ICE

John F. Vesecky, Martha P. Smith and Ramin Samadani

STAR Laboratory, Electrical Engineering Department
Stanford University, Stanford CA 94305-4055

ABSTRACT

The movement and characteristics of sea ice can be well observed by synthetic aperture radar (SAR) over large temporal and spatial scales. Since SAR observations can be made through clouds and fog and without solar illumination, they are well suited to polar remote sensing. We report image processing techniques for extracting the characteristics of pressure ridge features in SAR images of sea ice. The methods are applied to SAR images of the Beaufort Sea collected from SEASAT on October 3, 1978. Bright filamentary features are identified and broken into segments bounded by either junctions between linear features or ends of features. Ridge statistics are computed using the filamentary segment properties. Estimates of the density of sea ice ridging and the distribution of lengths and orientation are made. The information derived is useful in studying sea ice characteristics for ice science in remote sensing (ice classification) and in polar off shore operations (ship routing).

Keywords: SAR, Sea Ice, Pressure Ridges, Image Processing

I. INTRODUCTION

The movement and characteristics of sea ice can be well observed by synthetic aperture radar (SAR) over large temporal and spatial scales. SAR observations can be made through clouds and fog and without solar illumination. SAR observations of the polar regions are planned for the early to mid-1990's by the European Space Agency's ERS-1 satellite, the Japanese ERS-1 and the Canada/USA Radarsat. An important part of the ground segment of these polar observation programs is automated algorithms for image interpretation which will produce geophysical data products from the SAR observations. In this paper we present algorithms for extracting the characteristics of pressure ridge features in SAR images of sea ice. The algorithms are applied to SAR images of the Beaufort Sea collected from SEASAT on October 3, 1978.

Pressure ridges are thought to appear in SAR images as irregular bright filaments on ice floes. This interpretation of SAR sea ice images is made by analogy with visual and photographic observations of sea ice and on ice experience, e.g. see Welsh et al. (1986). Pressure ridge features in sea ice are important in both sea ice science and applications. Ridges structures are related to fracture by internal stresses, momentum transfer from wind to ice (drag coefficient), heat transfer and keel structures beneath the ice. Ridge characteristics are important for many sea remote sensing applications, such as ship routing, the safety of offshore structures and ocean acoustic wave scattering. SAR observations by satellite or aircraft coupled with automated image interpretation can provide accurate, consistent and timely information to both the sea ice science and applications communities.

Our objective in this brief paper is to summarize our work in automated extraction and characterization of pressure ridge features in SAR sea ice images. We consider a small region (12 x 12 km) of a SEASAT SAR image containing both leads and pressure ridges. A succession of image processing operators is applied to this image resulting in the extraction of ridge features. These features are then characterized in terms of ridge density (total ridge length per unit area), ridge length and ridge orientation. These statistics are then used to map ridge density over the image and to find the distribution of ridge lengths and orientations. None of these quantities is uniformly distributed.

II. APPROACH

We begin our approach with the premise that pressure ridges in sea ice fields correspond to bright filamentary features in SAR images. This premise is based on several arguments. First, bright filamentary features in SAR images correspond in form to pressure ridges observed in aerial photography. Second, the typical pressure ridge is composed of upraised blocks of fractured sea ice. Thus the ridge is typically much rougher than the surrounding sea ice on the scale of the radar wavelength (≈ 23 cm in SEASAT SAR images, such as Fig. 1). Further, in most situations, one face of the ridge is

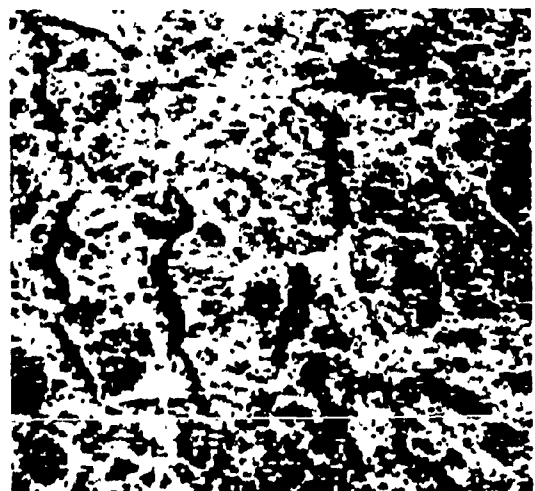


Fig. 1. SEASAT SAR Image of sea ice in the Beaufort Sea near 74° N, 125° W, collected on orbit 1409 on October 3, 1978. The pixel size is 100 x 100 m and the image area is 12.8 by 12.8 km. The SAR data were digitally imaged at Jet Propulsion Laboratory Nov. 19, 1986.

tilted toward the radar so that the radar observes it at nearer normal incidence. The tilted and relatively rough surface of the ridge backscatters radar waves more strongly at oblique incidence than would the horizontal and relatively smoother ice surrounding the ridge (Valenzuela, 1978).

Our approach to ridge extraction and characterization begins with two parallel operations, a thresholding operation to identify bright regions in the image and a line detection operation using a local Kasvand line detection operator. The results of these two operations are put through an 'and' operation to find linear features which are associated with bright regions. This operation identifies pixels which belong to pressure ridge features. To characterize ridge features we first thin the features to a width of one pixel, i.e. 100 m. We then find all the nodes of this thinned network of ridges, i.e. the intersections and end points. This operation breaks the ridge network into segments. We compute the ridge statistics using the properties of these segments, i.e. their density, length and orientation. These statistics are the raw data for sea ice science and applications as discussed below. This approach and the image processing techniques used to implement it are summarized in the block diagram of Fig. 2.

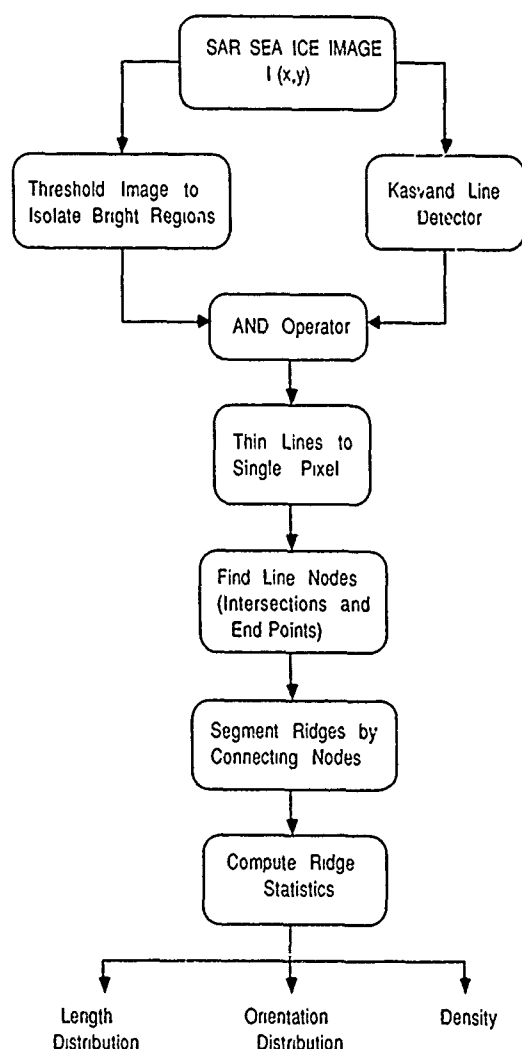


Fig. 2. Schematic block diagram of processing algorithm to extract and characterize pressure ridge features in SAR images.

III. IMAGE PROCESSING TECHNIQUES

The image processing techniques summarized in Fig. 2 begin with the raw image data of Fig. 1. This input image has been averaged to produce 100 x 100 m pixels. Note that there are filamentary structures associated with both lead-floe boundaries and pressure ridge features. To isolate pressure ridges we must do more than simply find linear features. The procedure used here is an 'and' operation. The inputs to the 'and' operation are bright regions found by thresholding the image to isolate the brightest 20% of the pixels. The result of the thresholding operation is shown in Fig. 3. Note the very confused structure. Fig. 3 shows that thresholding alone is not sufficient to isolate pressure ridge features.

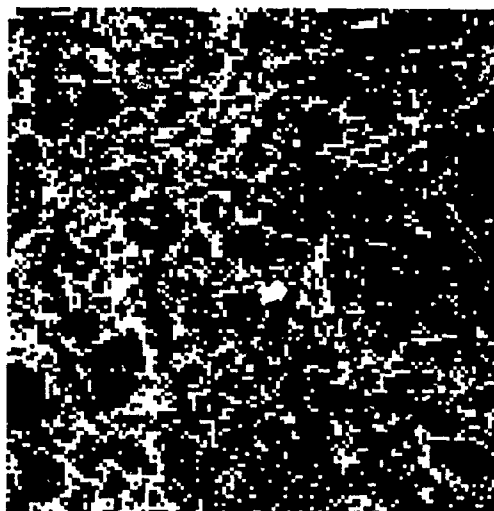


Fig. 3. Image of Fig. 1 processed by a threshold operation to show the brightest 20% of the pixels in white.

The filamentary features are found by applying a local line detector. The Kasvand line detector was implemented for this study. This method convolves a series of 5 x 5 pixel templates with the image pixels to find locally linear features. We are also studying more advanced filamentary feature detectors for application to speckled images (Samadani et al., 1989). The lines detected could be used as they stand or culled to remove short lines. In the study performed here we operated in two ways. For the main study we considered only those features with 6 or more pixels. We also maintained a data set without culling for comparison.

The next step is to combine the bright pixels with the pixels in linear features. This is done by performing a logical 'and' operation on the bright pixel image of Fig. 3 and the output of the line detector. The result of this operation is shown in Fig. 4. Note that filamentary ridge features are more prominent than in Fig. 3 where only brightness was considered. In addition the filamentary features constituting the lead boundaries in Fig. 1 are removed by our processing to find ridges.

In order to determine ridge statistics a better defined image is easier to work with. We therefore thin the filamentary features of Fig. 4 to a width of one pixel. This is done with the algorithm of Sakai et al. (1972) as implemented on the HIPS (Human information processing laboratory's Image Processing System) software package. The thinning operation applied to Fig. 4 results in Fig. 5. Some of the features now have less than 6 pixels due to the thinning operation.

We now deal with the problem of filamentary features that are interconnected. We determine all nodes, i.e. feature intersections and end points. This allows us to break up the multiply connected features and compute our statistics on the segments connecting nodes.

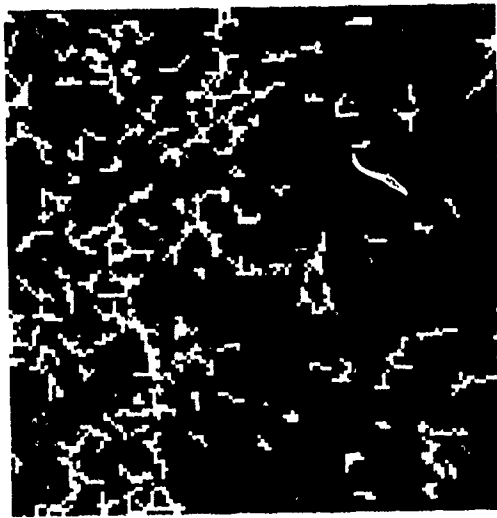


Fig. 4. Result of logical 'and' operation to determine pixels which are both bright and in a linear feature.

Several ridge statistics are computed here although other statistical quantities could be defined and computed from the information derived to this point. The ridge feature statistics computed here are the density of ridging per unit area, the distribution of ridge lengths and the distribution of ridge orientations. The density of ridges could be computed in several ways. We begin by defining the ridge density as the number of pixels in ridge features in a 5 x 5 km area. There are four such blocks in Fig. 1 with a little area left over. The features could be defined in several ways such as the bright pixels in the thresholded image of Fig. 3, the results of the logical 'and' operation shown in Fig. 4, or the thinned ridge features in Fig. 5.

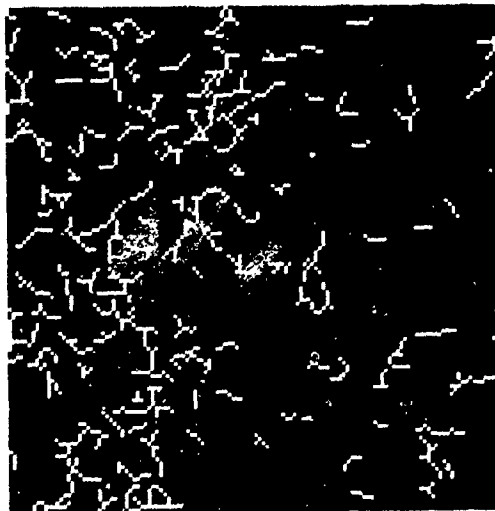


Fig. 5. Thinned ridges resulting from applying the thinning operation of Sakai et al (1972) to the results in Fig. 4. Features are restricted to a width of one pixel.

The length of a ridge segment is simply the total length along the filamentary feature between nodes. This may well be significantly longer than the distance between nodes. The orientation of a ridge segment is the orientation of the line between the nodes defining the segment.

IV. RIDGE STATISTICS RESULTS AND APPLICATIONS

The density results are shown schematically in Fig. 6. The numbers in the 5 x 5 km boxes represent different methods for estimating the ridge density per unit area. The top number uses the thresholded pixels of Fig. 3 to determine the ridge density. The two numbers in the left hand column use the results for ridge features having no limits on the number of pixels, i.e. ridges with < 6 pixels are included. The upper of these two numbers uses the unthinned ridges, while the bottom number uses the thinned ridges. The two numbers in the right hand column use only those features with 6 or more pixels, before thinning. The upper number uses the unthinned ridges of Fig. 4, while the bottom number uses the thinned ridges of Fig. 5. A comparison of the results of Fig. 6 with the original image of Fig. 1 shows that all these measures of ridge density are qualitatively correct. The region at lower left clearly has the most ridge features.

| | | | |
|-----|-----|-----|-----|
| 687 | | 522 | |
| 475 | 342 | 364 | 217 |
| 381 | 257 | 304 | 170 |
| 755 | | 476 | |
| 541 | 456 | 301 | 172 |
| 404 | 324 | 252 | 129 |

Fig. 6. Ridge density distribution for 5 x 5 km areas in Fig. 1. The thin line indicates the outline of the total area of Fig. 1. The numbers represent different methods of calculating the ridge density per unit area.

The ridge length distribution is shown in Fig. 7. The distribution is rather flat for the shorter lengths (< .6 km) and decreases markedly at the longer lengths. The average ridge length is some 400 to 600 m with few ridges longer than 1 km. A similar distribution was computed using all the features regardless of number of pixels. It shows more features, but is not significantly different in shape.

The orientation distribution is shown in Fig. 8. This distribution was calculated using only those features with more than 6 pixels before thinning. Again a distribution was also calculated using features of all lengths, and the distribution shape was not significantly affected. A distinct preference for the horizontal direction (0° or 180°) is shown in the distribution.

Histogram of Thinned Features

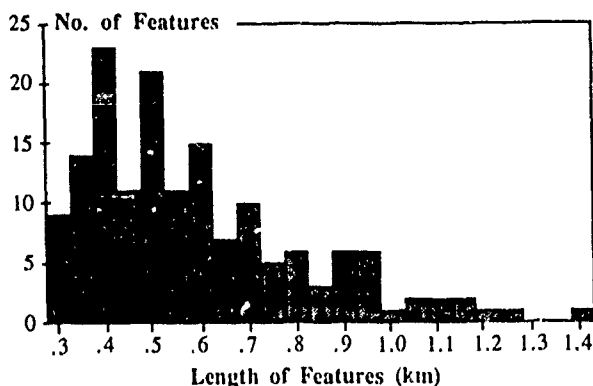


Fig. 7. Histogram of ridge lengths determined from data including only features of 6 or more pixels.

Histogram of Thinned Features

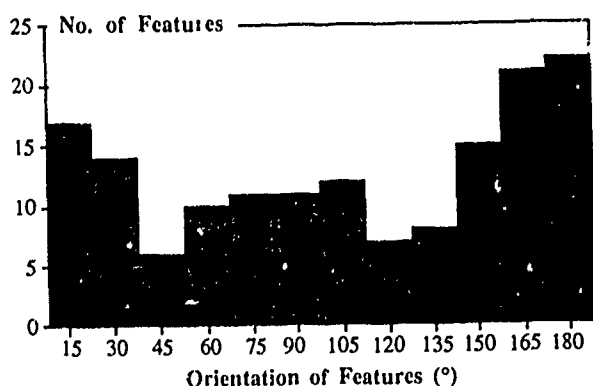


Fig. 8. Histogram of ridge orientations, determined from data including only ridge features of 6 or more pixels.

Applications for the information derived here occur in ice science and for off shore operations in polar regions. Ridge structures are important in determining air drag as discussed by Burns and Wegener (1988). In air drag calculations the ridge statistics considered here could form an input to estimate the drag coefficient. The statistical structure of the fracturing of the ice is indicative of internal and external stresses imposed upon the ice. Thus ridge measurements as demonstrated here could be used as information on the history of the sea ice observed and possibly as a predictor for future lead and ridge formation. Sea ice keels are the undersea counterpart of the ridges on the top of the ice. Thus the ridge information extracted here could be useful in estimating the keel structure underneath the ice.

For off shore operations in polar regions pressure ridges present a significant hazard. The ridges and corresponding keels are the thickest part of the ice and hence present the greatest hazard to even ice strengthened ships. Similarly ridges also form a significant hazard to ice off shore engineering structures.

V. CONCLUSIONS

We summarize our conclusions as follows:

1. Image processing techniques have been demonstrated to show that ridge structures in SAR sea ice images can be extracted automatically.
2. Ridge structure statistics can be estimated from SAR sea ice images.
3. Ridge structure information derived from SAR images should be useful in both ice science and off shore operations.

ACKNOWLEDGEMENTS

We are grateful to John Curlander and Ron Kwok at Jet Propulsion Laboratory for providing the SAR image used in this research. We also thank NASA Oceanic Processes (Bob Thomas) and ONR Remote Sensing (Chuck Luther) for both financial support and helpful suggestions regarding the research.

REFERENCES

- Burns, B. A. and A. Wegener, SAR image statistics related to atmospheric drag over sea ice, 409-412, *Proceedings of IGARSS '88 Symposium*, ESA SP-284, ESA Publications Div., Paris (1988).
- Richards, J. A., *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin (1986).
- Sakai, M. et al., *Comp. Graphics & Image Processing*, 81-96 (1972).
- Samadani, R. and J. F. Vesecky, Finding Curvilinear Features in Speckled Images, *Proc. IGARSS'89 Conference*, Univ. British Columbia, Vancouver, B.C. (1989).
- Valenzuela, G. R., Theories for the interaction of electromagnetic and oceanic waves--a review, *Boundary-Layer Meteorol.* 13, 61-85 (1978).
- Vesecky, J. F., et al., Observation of sea-ice dynamics using synthetic aperture radar images: Automated analysis, *IEEE Trans. Geosci. and Rem. Sensing*, GE-26, 38-48 (1988).
- Welsh, J. P. et al., A Compendium of Arctic Environmental Information, *NORDA Report 138*, NSTL, Mississippi (1986).

AUTOMATIC ESTIMATION OF ICE KINEMATICS USING REMOTE SENSING DATA

H.Flesche, K.Kloster, T.Olaussen, O.M.Johannessen

Nansen Remote Sensing Center

Edvard Griegs vei 3a, N-5037 Solheimsvik-Bergen, Norway

Telephone +47 5 29 72 88, telefax +47 5 20 00 50

Abstract.

This paper addresses automatic estimation of sea ice kinematics using a time sequence of satellite images. The aim of this study was to arrive at an efficient (non-calculation intensive), but at the same time accurate algorithm, based on the image pyramid method.

Some other algorithms also use the image pyramid method. A new addition in this approach is the search method. Instead of performing a total search in a limited area, a two-dimensional binary search is used. This can reduce the number of correlations by orders of magnitude, depending on the size of the search area.

The results from this system was compared with a manually estimated motion in a pair of satellite images. The two vector fields showed a high degree of correspondence. In addition to this test, synthetically constructed image sequences with shear motion were run through the system to get a quantitative measure of the errors in the vector field. These tests showed a percentage of correctly estimated vectors between 70% and 97%, depending on different factors that will be discussed.

1. Introduction.

In the last years there has been much activity in the field of ice motion estimation. Most of the methods are based on estimation of correlation factor between blocks in a pair of satellite images. The simplest way is to divide the images into square blocks and make a total search for the highest correlation factor in a limited area. The main drawback with this method is the amount of calculations that is needed to get a vector field of acceptable density.

Some methods has been suggested to lighten the burden of calculations. Collins & Emery [2] and Ninnis et. al. [8] use a Fourier transform of the images for the motion estimation (Matched filtering). By calculating the correlation in the frequency domain instead of the time domain the convolution between the image blocks is transformed to a multiplication, and the motion can be estimated in one operation.

The approach to the problem that this work is based on is the image pyramid method. Earlier work on

this method has been done by Fily & Rothrock [3] and Samadani [11].

The algorithm is planned to be used to extract ice-kinematic information from the satellite imagery that will be acquired by the ERS-1 satellite. The results of the motion estimation will be used as part of the input to a Hibler-Preller dynamic/thermodynamic model on the Barents Sea that is developed at NRSC in cooperation with NORDA. For references, see [5], [9] and [10].

2. Image Pyramid Method.

An image pyramid is a set of images derived from a single image, all showing the same motive but with varying resolutions. The pyramid is made by low-pass filtering and sub-sampling the original image. The resultant image is shown in figure 1. The low-pass filter used in this study is developed by Burt [1].



Figure 1: Image Pyramid.

The advantage of the image pyramid is that a rough estimate of the motion can be calculated from the images with the coarsest resolution with few correlations. This estimate will be used as the start value for a search on the next level in the pyramid. Due to this start value we can limit the search on the higher levels to a smaller area.

The image pyramid method also has an advantage in its possibility of noise reduction. This can be done during the motion estimation by filtering the vector field. A common method is median filtering. Median filtering will remove noise spikes, i.e. vectors that are unlikely compared with their neighbors. Several methods of median filtering has been evaluated, and the best results up to now have come from an IIR median filter that filters the displacement in the two directions separately.

The use of median filtering is based on the assumption that the displacement vector field will be consistent. Since the motion of the ice floes and pack ice in adjacent image blocks cannot be totally inconsistent, this is a natural assumption to make.

For a more exhaustive description of the image pyramid method, see [11] and [12].

3. Binary Search.

The new part of this system is the search algorithm. This is a two-dimensional extension of the binary or logarithmic search. This method was developed by Jain & Jain [6] and modified by Kappagantula and Rao [7]. It was originally intended for motion compensated coding of video signals, a field closely related to this motion estimation.

The one-dimensional binary search requires a sorted data set and a distance measure that is monotonically increasing away from the optimal position. This requirement must also be fulfilled in the two-dimensional case for both dimensions if the algorithm shall be able to find the optimal displacement. The distance (or rather closeness) measure as employed here is a normalized correlation coefficient. Tests show that this measure will be monotonically decreasing in both dimensions within a limited area around the optimal position. The extent of this area depends on the distribution of spatial frequencies in the image and the window size of the correlation function. Therefore, if the search shall succeed, the peak of the correlation function must be included in the initial search area, see fig. 2 & 3.

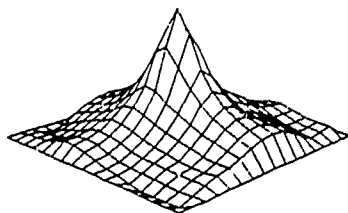


Figure 2: The peak of the correlation function. The function is monotonically decreasing in both dimensions.

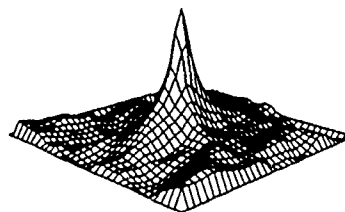


Figure 3: The correlation function calculated over a larger area. The function is not monotonically decreasing.

The algorithm can be described as follows:

1. Given a sub-block of image 1 and a search area in image 2.
2. Start the search by calculating the correlation that corresponds to zero displacement. Then find the correlation for four areas, the position of which corresponds to a displacement d that is half of the maximum displacement along the axes from the origin. If the center position has the highest correlation go to step 3b, otherwise go to step 3a.
- 3a. Find the correlation for two areas located at a distance d normal to the direction from the origin to the last optimal position. Compare these and the last optimal position and use the best as the new starting point in step 3b.
- 3b. Do the same as in step 2, but with $d = \lceil d/2 \rceil$.
4. Continue the search until $d < 1$.

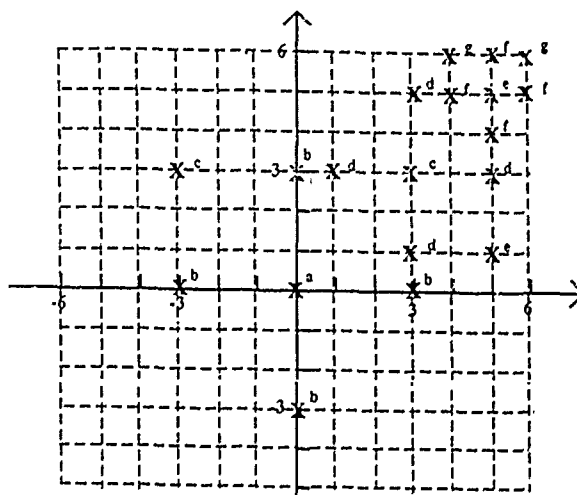


Figure 4: Search pattern with a maximum displacement of 6 pixels.

To clarify this we show an example of the search with maximum displacement = 6 pixels in figure 4. The search starts in the origin (a) which corresponds to zero displacement. The correlation coefficient is compared with four other values as described above. The position of these is marked b in fig. 4. Position (0,3) is the optimal (marked with a grey X). According to step 3a above two positions marked c are compared with the last value. From this comparison the position (3,3) is the optimal. The search goes on with $d=2$ to the positions marked d and e, and at last with $d=1$ to positions f and g where the final optimal position within the search area is found.

4. Results.

The testing of the algorithm was done with a pair of NOAA images from the Fram Strait and with pairs of images that had a synthetically generated motion. The latter part of the test was performed to obtain a quantitative error measurement.

4.1. Testing with a synthetic motion.

Two pairs of images were used. Both had a shear motion with a displacement of 5 and -5 pixels on the image halves. One pair had a stretched histogram and thereby a good contrast, see fig. 5, while the other pair had a very narrow histogram, see fig. 6. The results from the tests are given in tables 1 and 2 and an example of displacement vectors in figure 7.

| Window size | %correctly estimated |
|-------------|----------------------|
| 8x8 | 71 |
| 16x16 | 88 |
| 32x32 | 97 |
| 64x64 | 96 |

Table 1: Results from image pair with stretched histogram.

| Window size | %correctly estimated |
|-------------|----------------------|
| 32x32 | 78 |
| 64x64 | 79 |

Table 2: Results from image pair with narrow histogram.

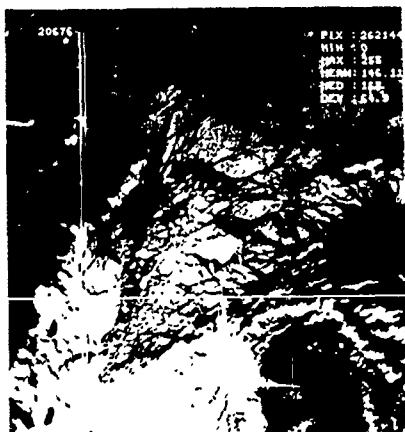


Figure 5: Image with stretched histogram.



Figure 6: Image with narrow histogram.

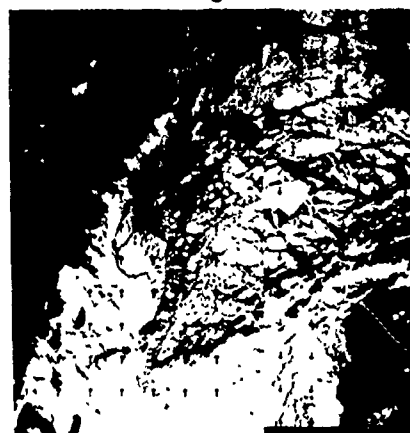


Figure 7: Displacement vectors.

As can be seen from tables 1 and 2 the contrast plays an important role for the result. Stretching of the histogram should therefore be done as a preprocessing of the image pair.

In addition to this we see that the size of the window has an impact on the results. A large window will produce a more confident correlation estimate. Opposed to this is the system's ability to describe motions that are not purely translational. These are estimated more correctly if the window size is smaller and the vector field more dense. Judging from table 1 a window size of 32x32 pixels seem optimal. This parameter must probably be adjusted according to the satellite data used.

4.2. Testing with a real sequence of images.

The purpose of the testing done with the real sequence of images was to compare manual and automated vectors.

In most part of the images the automated vectors correspond to the manual ones. Problems arise in areas which are cloud covered in one or both of the images. These vectors have a low correlation coefficient and can be pruned by setting a threshold on the correlation coefficient. Featureless areas also give a poor result for both manual and automated vectors. The best results are achieved in the central part of the image, where there

are several distinct ice floes. By superimposing the two time-sequence images one can confirm the accuracy of the extracted displacement vectors. Figures 8 and 9 show the automated vectors overlaid on the image sequence.



Figure 8: Displacement vectors overlaid on the first image in the sequence.

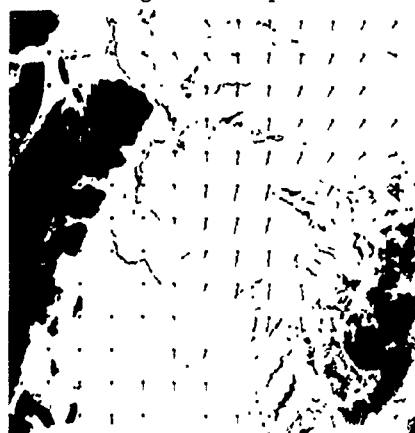


Figure 9: Displacement vectors overlaid on the last image in the sequence.

5. Conclusions.

In this paper we have discussed a fast algorithm for sea ice motion estimation.

We have shown that the algorithm gives an acceptable result with a real sequence of satellite images and an error rate of 3-30% with a synthetic motion.

The result from the algorithm depend on several factors. We have shown that the window size the maximum displacement and the noise reduction system (i.e. the median filtering) all have impact on the result. The parameters for window size and maximum displacement should be adjusted to fit the satellite system that is employed.

This system was developed for image processing systems with little capacity, and therefore emphasis was put on the effectivity of the algorithm. Using the 2D-binary search algorithm the motion estimation on a pair of 512x512 pixels images takes 27 minutes with a Sun 3E processor (similar to Sun 3/60 workstation).

The search algorithm will probably not be able to give a good motion estimate under all conditions. It is dependent on a monotonically decreasing correlation function within the search area in order to be able to detect the optimal position. An image that generally has small features and thereby high spatial frequencies, will have a smaller area where the correlation function is monotonically decreasing. The results are being evaluated in order to develop a more comprehensive algorithm that can perform motion estimation under more variable conditions.

Acknowledgement.

I would like to thank professor Tor Ramstad and Jon Håkon Husøy at the Norwegian Institute for Technology for their help with my M.S.E.E. thesis which is the main part of this work.

H. Flesche

References.

- [1]: Peter J. Burt: Fast filter transforms for image processing. *Computer graphics and image processing*, vol 16, pp 20-51 (1981).
- [2]: Michael J. Collins & William J. Emery: A computational method for estimating sea ice motion in sequential Seasat synthetic aperture radar imagery by matched filtering. *Journal of geophysical research*, vol 93 no c8, pp 9241-9251. August 1988.
- [3]: Michel Fily & D. A. Rothrock: Sea ice tracking by nested correlations. *IEEE transactions on geoscience and remote sensing*, vol GE-25 no 5, pp 570-580. September 1987.
- [4]: Harald Flesche: Estimering av isbevegelser fra satellittbilder (Estimation of ice motion from satellite images). M. Sc. thesis at the Norwegian Institute for Technology, Dep. for electrical and computer engineering. December 1988. (Norwegian)
- [5]: W.D. Hibler III: A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography* 9, 1979, pp 815-846.
- [6]: Jaswant R. Jain & Anil K. Jain: Displacement measurement and its application in interframe coding. *IEEE transactions on communications*, vol COM-29 no 12, pp 1799-1808. December 1981.
- [7]: S. Kappagantula & K.R. Rao: Motion compensated interframe image prediction. *IEEE transactions on communication*, vol COM-33 no 9, pp 1011-1014. September 1985.
- [8]: R. M. Ninnis, W. J. Emery & M. J. Collins: Automated extraction of pack sea ice motion from AVHRR imagery. *Journal of geophysical research*, vol 91 no c9, pp 10725-10734. September 1986.
- [9]: T.Olaussen, P.M. Haugan & R.H. Preller: Integrated remote sensing for Barents Sea ice, monitoring and modelling. *Proceedings from Polar Tech'88*, Trondheim, Norway, pp 485-506.
- [10]: R.H. Preller: The NORDA/FNOC polar ice prediction system PIPS- Arctic: A technical description. *Naval ocean research and development activity, NSTL, MS. 1585*, 63pp.
- [11]: Ramin Samadani: Image pyramid motion detection applied to sea ice. Ph. d. thesis at CASIS, Dep. for electrical engineering, Stanford University. September 1987.
- [12]: John F. Vesecky, Ramin Samadani, Martha P. Smith, Jason M. Daida & Ronald N. Bracewell: Observations of sea ice dynamics using synthetic aperture radar images: automated analysis. *IEEE transactions on geoscience and remote sensing*, vol 26 no 1, pp 38-48. January 1988.

IGARSS'89 ABSTRACT

SUGGESTED SESSION: F. Ice/Oceans/Air: Polar Ocean Research

ANALYSIS OF ICE MOTION VECTORS DETERMINED BY THE ASF GPS
FOR SEASAT SAR OBSERVATIONS IN THE BEAUFORT SEA

By Frank Carsey, Benjamin Holt and Ronald Kwok,
Jet Propulsion Laboratory, Pasadena CA 91109 USA,
Telex 675-429, FAX 818 354 3437, Ph 818 354 8163

The Alaska SAR Facility (ASF) is a NASA program to implement and operate a facility at the University of Alaska Fairbanks for the acquisition and processing of data from ERS-1, and to administer science program investigations using the data. The original configuration of ASF consisted of the Receiving Ground Station, the SAR Processing System and the Archive and Operations System (AOS). These major systems making up the ASF are to be augmented by image analysis processors designed to develop specific geophysical information from the SAR images. A Geophysical Processing System (GPS) is being implemented which will routinely process SAR images of sea ice to produce fields of ice motion and ice-type concentration to support studies of ice dynamics and to update operational predictive ice models for the ice covered waters in the ASF station mask. To accomplish the determination of ice motion and ice type the GPS will utilize a complex ice segmentation and tracking algorithm. This procedure will: Segment and classify the ice floes in a newly acquired image, predict the previous trajectory of the floes in the image, search the AOS for a suitable previous image along the predicted trajectory, ingest the previous image, perform segmentation on the previous image, determine ice motion through a hierarchical tracking analysis on the image pair, interpret the quality and comprehensiveness of the tracking procedure (relative to whether it should be repeated on another previous image), and write the derived information back into the AOS. For this study the GPS in its interim configuration was used to process a sequence of SEASAT SAR images taken in the eastern Canadian Basin about 500 km west of Banks Island. The objectives of the study are to examine the properties of the ice deformation in relationship to the source scale of the relative velocity field, and to verify that the GPS is capable of generating reliable ice motion data under various conditions of rotation and translation. The images were selected to follow an (unnamed) ice island fragment which was previously studied in the SEASAT record by Fu and Holt and by Arsenault. The SAR record begins in the summer and continues through to the fall of 1978, and thus the data set is approximately Lagrangian; it contains a largely fixed collection of floes as they move under wind and current forcing and undergo seasonal change. The derived ice motions are analyzed for kinetic variables and for the behavior of the feature spacing, D , with time. The behavior of the square of the time derivative of D is compared to predictions that it will behave according to a power law with an exponent that relates the source of the motion to a mix of local and large-scale forcing.

PRELIMINARY OBSERVATIONS OF LABRADOR SEA MARGINAL ICE ZONE RHEOLOGY USING C-BAND SAR

Mark R. Drinkwater
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive, Pasadena
CA 91109, USA

Vernon A. Squire
Department of Mathematics and Statistics
University of Otago
PO Box 56, Dunedin
New Zealand

ABSTRACT

Synthetic aperture radar imagery collected in the Labrador Sea during LIMEX '87 are used to interpret modes of sea ice deformation. The ice canopy exhibited two distinct rheologies separated by a clear line of shear; a quasi-brittle inner regime and a non-linear viscous outer regime. A single constitutive relation capable of modelling both is unlikely within a plastic rate-independent formulation. Rate dependent effects are discussed as an explanation for brittle fracture in ductile materials.

Keywords: LIMEX, rheology, plastic, brittle, ductile.

1. INTRODUCTION

Few studies of the cryosphere have utilised SAR imagery directly for a geophysical purpose, undoubtedly due to the complexity and detail of these data. The exception is work to determine sea ice displacement, which has proven the potential of SAR in mapping sea ice motion and deformation fields with high precision and resolution. In this paper we extract kinematical information from aircraft SAR data, and make suggestions regarding mesoscale ice rheology? An understanding of ice deformation is a prerequisite for sea ice dynamics studies. Radar images, give an appreciation of deformation and flow mechanisms, and are a hitherto uninvestigated source of rheological information.

Data were acquired by the Canada Centre for Remote Sensing (CCRS) 5.3 GHz aircraft SAR system during the March 1987 Labrador Ice Margin Experiment (LIMEX'87) [1,2]. The SAR system acquired C-band images on several flights over the Labrador marginal ice zone (MIZ). Single SAR images provide a qualitative picture of ice rheology, because as a two-dimensional continuum the ice field can only deform, yield or fracture in well-defined and well-documented ways [3,4]. Overlapping image pairs, separated by 1 or 2 days, enable quantitative information on deformation in the pack to be extracted. The resolution of the images is sufficient to study many features in the sea ice field that suggest a rheological scenario; the presence of a shear zone, for example, can immediately dismiss a simple linearly viscous model.

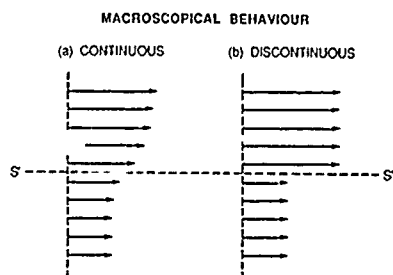


Figure 1 Displacement in laminar shear flow, a) continuous, b) discontinuous

At low strain rates and scales large compared with the characteristic dimensions of the "granular" structure of pack ice, the slow deformational response of the MIZ may be adequately described by plastic shear flow. This type of flow is conceptualised as a single set of infinitesimally close laminae sliding parallel to a shear line (Fig. 1a), or alternatively it may be inhomogeneous and display distinct discontinuities in the displacement along shear lines (Fig. 1b). Discontinuous shear flow leads to large relative displacements between parallel laminae. In fully developed shear flow, deformation does not occur until the stress reaches a critical yield level, and it begins and ends without appreciable delay as the stress is first applied and then reduced to a sub-critical value. All mechanical energy expended during the flow is dissipated and when fluidity ceases no mechanical energy is recovered. The SAR imagery reveals several zones within the Labrador Sea MIZ which display some of these characteristics; these are investigated from a rheological standpoint.

2. MARCH 1987 GRAND BANKS ICE CONDITIONS

During a normal spring, the Grand Banks region is a classical MIZ with a diffuse pack and seasonally varying ice edge. Meteorological and oceanographic influences strongly modulate the daily extent of the sea ice cover. MIZ morphology varies on all time and spatial scales due to local interactions and forcing by, for example, swell and storm-generated waves which break up the ice canopy, or winds and currents which cause ridging and over-raftering of floes. The tongue of ice that stretches down the Newfoundland coast and out over the Grand Banks in winter consists largely of moderately deformed ice less than 1.5m thick.

A rapid shoreward ice edge retreat took place from 11-25 March, 1987 in response to north-northeasterly winds, rather than melt back by persistent warm southerlies. Particularly vigorous ice compaction took place from 16-20 March throughout a spell of southeasterlies, during the east-west passage of two low pressure cells to the south of Newfoundland. Onshore winds acted to increase the shear strength of the ice cover through ice compaction, thereby leading to smaller ice velocities closer to shore.

3. LABRADOR MIZ KINEMATICS

3.1 Data Description

SAR images acquired between 17 and 26 March enabled a kinematical study of ice motion to be carried out using image pairs with spatially correlated features separated in time. The locations of images used in this analysis are shown in Fig. 2a. Spatial overlap was achieved in the Cape Freels region, south Bonavista Bay, and the northern area of Conception Bay, permitting coverage of both inner and outer MIZ.

3.2 Motion analysis

Ice motion was derived by plotting trajectories of recognizable features for the duration of available overlapping images. Velocity vectors are calculated from their displacement relative to coastal fixed reference

points. In most cases recognizable features included distinctive ice floe shapes or tonal patterns. Fig. 2b illustrates the regional pattern of motion vectors and velocities derived from motion pairs spanning the period bracketed by the images.

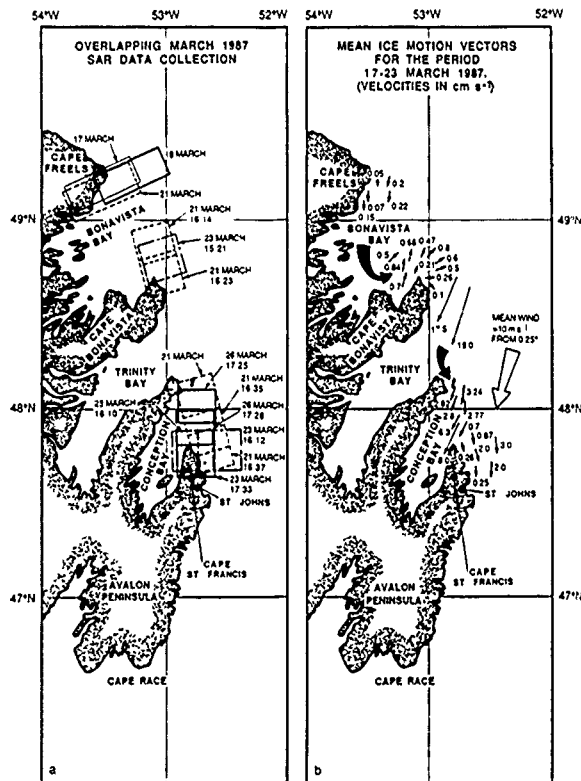


Figure 2 SAR data used for kinematic analysis a) locations of overlapping images taken at different time intervals, b) Ice motion field derived from the SAR images.

Ice advection in the more mobile outer MIZ is toward the south, in response to forcing by the Avalon channel filament of the Labrador current. Mean current speeds in this region are calculated from drift data collected during 1985-1987 to be between 7 and 10 cm s⁻¹ [5]. This is in agreement with surface current measurements made during LIMEX'87 [6]. Detailed analysis reveals a velocity gradient with distance from the shore. More significant, however, is the marked increase in velocities observed at the boundary between inshore ice and the slurry of wave-pomelled outer ice. This transitional region within the MIZ is identified as a major line of shear within the pack [7] (cf. Fig. 1). To the east of the line ice floes travel far more rapidly, at rates more commonly associated with northern sea ice moving along the edge of the Labrador shelf. Speeds up to 18 cm s⁻¹ are seen off Cape Bonavista, comparable with observations of ice motion along the Newfoundland shelf obtained with floe mounted beacons [5]. Importantly, the east-west shear transition becomes better defined between 21 and 23 March when winds shifted from southeasterly to north-northeasterly, thereby relieving ice pressure.

In some near-shore areas, there is a suggestion of topographic steering of currents or circulatory motions in some anomalous ice floe trajectories. For example, ice appears to enter Bonavista Bay around Cape Freels, and to leave around Cape Bonavista following the 200m isobath (Fig. 2b). In-flow of ice appears to occur around Cape Freels, with the ice flowing cyclonically around the Bay and then exiting northeastwards around the northern tip of Cape Bonavista. Though consistent with buoy trajectories [5], this pattern is inconsistent with increased ice pressure from the overall westward wind-induced retreat of the ice edge and the general southerly flux of ice. In the shelf region, the dominant tidal constituent has an amplitude of around 2-10 cm s⁻¹ which may influence ice motion on shorter timescales.

4. ICE RHEOLOGY

To date, most ideas on the rheological aspects of ice dynamics and basin-scale sea ice models have been based on *in situ* observations of sea ice structure and physical properties, and visual interpretation of surface expressions of deformation such as ridging, rafting or shearing. The use of SAR to infer deformational behaviour in seasonal pack regions is a promising new alternative to previous methods.

4.1 Introduction

Fig. 3 shows characteristic rheological responses of many materials in terms of stress and strain. Curves in Fig. 3a are applicable to both tension and compression. The initial linear behaviour of the stress σ with strain ϵ in the figure corresponds to elastic deformation. At point A (the yield stress), the elastic limit is reached and the material begins to yield by deforming plastically. In the ductile regime, from point A to point B, the plotted curve illustrates strain hardening, a property typical of many materials whereby the material strengthens with application of stress such that increased stress is required to increase the strain. This is in contrast to perfect plasticity where AB would be a line parallel to the ϵ axis. At point B on the curve a sample in tension in the laboratory would begin to neck, i.e., the area reduction in some region of the sample would not be compensated by the added strengthening effect of the plastic deformation. Soon after reaching point B the material fails and fractures.

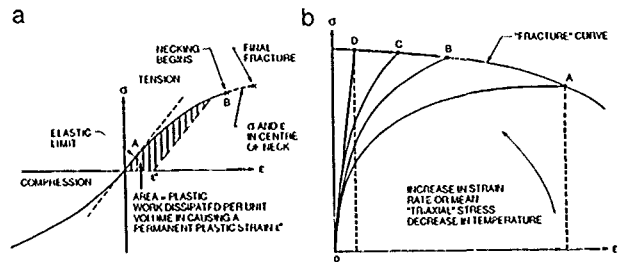


Figure 3. a) True stress σ versus true strain ϵ indicating elastic, plastic, and failure states. A is the yield stress which marks the onset of strain hardening, at B the material begins to neck. b) family of stress-strain curves showing the influence of factors which promote brittle fracture of a normally ductile solid.

Not all materials initially deform elastically followed yield and irreversible plastic deformation, both being time-independent. In many materials behaviour is time-dependent. Ductility can be affected by strain rate, and brittle fracture is encouraged by factors which tend to raise the stress-strain curve. In Fig. 3b a bounding fracture curve is drawn to intersect the various stress-strain curves representing tensile experiments done at lower temperatures, higher strain rates or increased stress states (in the ABCD direction). The total energy to fracture, i.e., the area beneath the curve, for a material deforming at high strain rates, for example, will be much less than that for the same material deforming at low strain rates. High strain rates favour brittle behaviour rather than ductility.

4.2 The Labrador MIZ

In spring 1987, the wave-fractured sea ice was a granular aggregate of various floe sizes and thicknesses. From a physical standpoint it is a continuum deforming under the action of forces that induce movements and irrecoverable changes in the ice canopy. Deformation or flow is controlled by the rheology of the continuum, and this is expressed mathematically as a constitutive equation relating stress and strain rate. A number of isotropic constitutive equations representing rate-independent visco-plastic ice deformation have been proposed, these are generalized by an expression of the form [8].

$$\sigma_{ij} = 2\eta d_{ij} + [(\zeta - \eta) d_{kk} - P^*/2] \delta_{ij} \quad (1)$$

where σ_{ij} and d_{ij} are the two-dimensional stress and strain rate tensors respectively, and δ_{ij} is the Kronecker delta. In the most general form of (1) the shear viscosity η , the bulk viscosity ζ and the ice strength P^* vary spatially and are functions of the strain rate invariants d_{kk} and $d_{ij}d_{ij}$. Equation (1) is a two-dimensional Reiner-Rivlin fluid, which

with simplification represents a variety of rheological behaviours, e.g., Newtonian viscous fluid, viscous-plastic fluid, etc.

There are compelling indications that any reduction of (1) to a linear form will not suffice at these spatial and temporal scales, and that a fully non-linear constitutive relation is needed if features of the Labrador MIZ are to be adequately modelled. A simple linear Newtonian fluid with no bulk viscosity, for example, will not resist convergence. The inclusion of bulk viscosity to provide compressive strength overcomes this deficiency, but requires different viscosities near to, and far from the shore. Non-linear plastic theories avoid these problems and have been especially successful in modelling the ice of the Arctic Basin. Such an ice cover has low tensile but high compressive strength. A plastic constitutive law allows these attributes to be ascribed simply by specification of the yield curve, but behaviour before yield and the precise form of the yield curve, and hence the form of η , ζ and P^* , are subjects of debate. Various methods have been used to account for strain hardening (or softening), by making P^* a function of ice thickness and compactness [8,9].

4.3 Balance of Forces

The balance of forces acting upon interior ice off the Avalon Peninsula are estimated for two periods of time when wind conditions were reasonably stationary for long periods. Estimates of air-ice and ice-water stress are based upon measured drag coefficients of ice with similar properties to Labrador Sea ice and wind and current data. If forces upon a floe are assumed in equilibrium, then wind and current stresses when added to the Coriolis force F_c should be balanced by the internal ice stress F_{int} . Pressure gradients generated by tilt of the ocean are assumed negligible. Thus, for ice floes far enough into the MIZ to be unaffected by wave radiation stresses, the balance in Fig. 4a is assumed from 19-21 March. Ice velocity is included, though it does not enter into the balance. Another force balance is constructed for the period 21-23 March, in Fig. 4b, after winds shifted to north-northeasterly.

Internal ice stress responds according to the shift in wind direction on 21 March and there is a 90° rotation of principal stresses to a northeast-southwest direction. Floe trajectories off the northern tip of Cape St. Francis (Fig. 5) match this direction, and ice velocity V_i increases to a maximum of 8 cm s⁻¹. Importantly, this change in the combination of wind and current-induced stress appears related to initiation of brittle shear fracture in the ice canopy. Velocity vectors become rotated through an angle of 10° from their previous direction under the influence of a greater wind stress, tracing the shear line.

4.4 Observed Rheological Regimes

Deformation of the ice canopy appears dependent on geographical location, due to variations in the ice-water mixture and the inhibiting influence of a landmass. Suggestions of both brittle failure and plastic flow are evident and Fig. 5 shows a region off the northern tip of the Avalon peninsula. Recognizable floes are plotted on 21 and 23 March in Fig. 5a, and b, indicating motion in the pack across a shear line. Maximum relative shear motion occurs between 21 and 23 March, and by 26 March swell had penetrated and motion ceased. Between 21 and 23 March the maximum velocity differential observed across the shear line is about 7 cm s⁻¹ and the velocity gradient indicates a considerable mean shear strain rate of 1.24 d⁻¹. Most of the deformation occurs in the plane of the shear line, there are few extensional faults suggesting that tensile stresses did not play a part in the distortion. The mode of failure illustrated in Fig. 5 resembles simple shear fracturing. Under biaxial stress, fracture lines can form at angles of less than 45° with the direction of maximum compression.

The shear line in Fig. 5 may be compared to strike-slip faults seen in horizontally-stressed rock formations [3]. Although such features can also be generated by a rate-independent plastic rheology, substantially more work is required to create the same effect than can be achieved given the forces in Fig. 4. This behaviour suggests brittle failure at high strain rates (as this requires the least energy to fracture) leading to discontinuous deformation of the compacted ice field. The association of high strain rates and fracture observed during LIMEX '87 therefore conflict with the notion of a rate-independent rheology for the MIZ. Rate-dependent effects are important, as increasing the strain rate encourages brittle fracture at the expense of ductility.

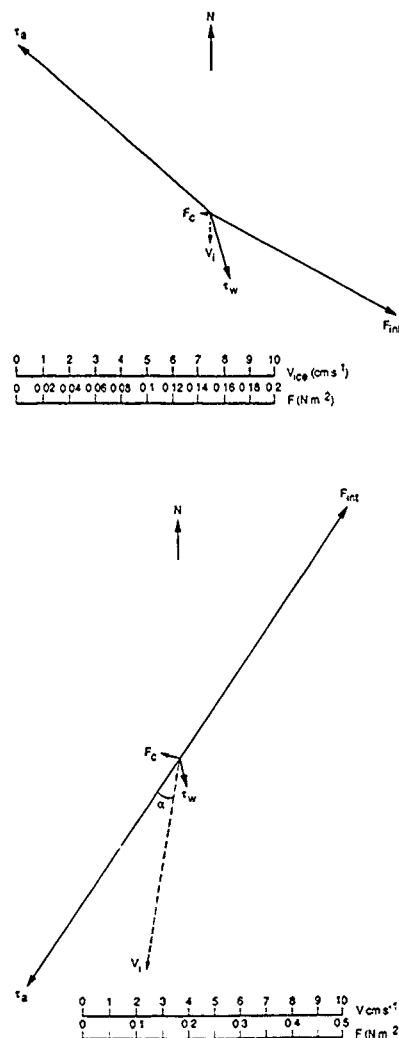


Figure 4. Balance of forces on ice floes off the tip of Cape St. Francis during two 3 day periods in March 1987, a) under southeasterly winds between 18 - 21 March, and b) after winds shifted to north-northeasterlies from 21 - 23 March. The ice velocity vector V_i is indicated.

The outer MIZ has a significantly different floe size distribution and concentration to the interior [10], and deforms continuously under conditions more typical of viscous, but not Newtonian, flow. A bulk viscosity can be included to allow the constitutive equation to respond to changes in the ice strength induced by factors such as wind, current and wave-induced radiation stress at the edge. Surface observations suggest that the medium flows under any significant applied deviatoric stress, however small, suffering continuous change with time. Fig. 6 indicates the differences between the outer MIZ flow regime and the inner MIZ brittle regime. Bright bands of small floes have migrated into the ice slurry from the edge and the effect of an eddy is passively traced by the more ductile ice. Ductility in a material is known to be affected by strain rate and the mean rate of strain observed perpendicular to the edge is in excess of 0.62 d⁻¹. Streamlines within the ice indicate that the outermost floes are being influenced by both eddy vorticity and an ice edge jet, enabling the ice eddy to be necked off. The eddy feature is embedded in a field of ice tracing a cyclonic motion, while the MIZ edge jet exports remaining floes southwards.

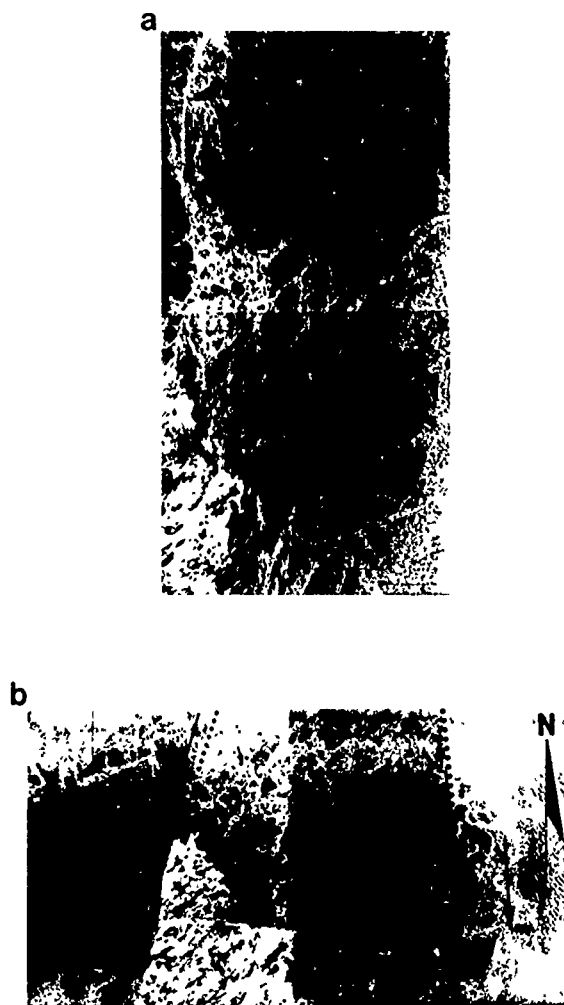


Figure 5. SAR mosaics constructed from overlapping scenes taken of sea ice off the northern tip of Cape St. Francis on: a) 21 March, b) 23 March. Several floes are identified in order to track deformation, and arrows indicate motion direction. Dotted lines indicate major shear zones. Images were provided by CCRS.

5. DISCUSSION

The rheological behaviour of sea ice is largely dependent on the level of applied stress and on scale. Rate-dependent effects appear important in some instances, especially in the premature initiation of brittle fracture. This detail appears to have been largely ignored in the ice modelling literature. Rate-independent rheologies can model dislocations of the kind seen in the LIMEX '87 data [26], but require more energy to initiate failure. The stress level and strain rate independence of plastic rheologies can be argued for continuous ice covers, where work goes into building ridges in a rate-independent manner, but this is less clear in the MIZ where strain rates can be much greater. Using the SAR images, strain rates can be deduced from the kinematic field in order to look at the relationship and concurrence between high strain rates and fracturing.

Two dominant rheological regimes existed during LIMEX '87. an inner zone experiencing rupture by brittle fracture, and an outer zone behaving as a non-linear viscous fluid. The shear boundary between the two regimes appears to be determined by the combination of principal stresses, and their relative dominance. It was observed to be either continuous (as on 21 March) or discontinuous (as on 23 March).



Figure 6. CCRS C-band SAR image of the outer MIZ on 21 March indicating bright bands of small floes tracing eddy vorticity in the outer edge, some of which have been dislocated by shear activity.

There are several advantages to using high resolution SAR imagery to study sea ice rheology. Notably, time-sequential data enables deformation in sea ice regions to be studied directly, and fluxes to be quantified. Such investigations can map the kinematics of ice fields and, through associated oceanographic and meteorological data, quantify the forces that produce the deformation. The use of SAR data, whether from satellite or aircraft, will undoubtedly revolutionize our understanding of the rheology of both MIZ and interior sea ice.

6. ACKNOWLEDGEMENT

Thanks go to S. Paterson and S. Argus of CCRS/Intera for imagery, and personnel from the DAD and Airborne Operations Division at CCRS who contributed to LIMEX '87. Thanks are extended to others whose data have allowed this study to be undertaken. MRD did this work at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration, while supported by a National Research Council Resident Research Associateship. VAS acknowledges the continued support of the University of Otago and the University Grants Committee of New Zealand.

7. REFERENCES

- [1] L. McNutt, S. Argus, F. Carsey, B. Holt, J. Crawford, C. Tang, A. Gray, and C. Livingstone, "LIMEX '87 The Labrador Ice Margin Experiment, March 1987," *EOS*, vol. 69, no. 23, p. 634, 1988.
- [2] M.R. Drinkwater and S. Digby-Argus, "LIMEX '87: An international experiment in the Labrador Sea marginal ice zone," *Polar Record*. In press.
- [3] J.R. Marko and R.E. Thomson, "Rectilinear leads and internal motions in the ice pack of the Western Arctic Ocean," *J. Geophys. Res.*, vol. 82, no. 6, pp. 979-987, 1977.
- [4] B. Erlingsson, "Two-dimensional deformation patterns in sea ice," *J. Glaciol.*, vol. 34, no. 118, pp. 301-308, 1988.
- [5] J.K. Peterson, and G. Symonds, "Ice floe trajectories off Labrador and Eastern Newfoundland: 1985-1987," *Canadian Technical Report of Hydrography and Ocean Sciences*, no. 104, 101p., 1988.
- [6] Tang, C.L., "Oceanographic observations during LIMEX," *Canadian Data Rep. Hydrogr.*, Ocean Sci., 1988.
- [7] S. Digby-Argus, F. Carsey, and B. Holt, "Wind, current and swell influences on the ice extent and flux in the Grand Banks-Labrador Sea area as observed in the LIMEX '87 Experiment," *Proc. 2nd Conf. Polar Meteorol. and Oceanography*, (Madison, March 29-31), pp. 11-14, 1988.
- [8] M. Leppäranta and W.D. Hibler, "The role of plastic ice interaction in marginal ice zone dynamics," *J. Geophys. Res.*, vol. 90, no. C6, pp. 11859-11909, 1985.
- [9] J.E. Overland and C.H. Pease, "Modeling ice dynamics of coastal seas," *J. Geophys. Res.*, vol. 93, no. C12, pp. 15619-15637, 1988.
- [10] S. Digby-Argus and F.D. Carsey, "SAR imagery of the Grand Banks pack ice and its relationship to surface features," *Proc. IGARSS'88 Symp.* (Edinburgh, Sept. 12-16, 1988). In: *ESA SP-284*, vol. 3, pp. 1425-1428, 1988.

EVALUATION OF ICE CONDITIONS IN THE NORTHERNBERLAND STRAIT USING C AND X BAND SAR IMAGERY

J.A. Dechka, V.L. Shaw, F.G. Bercha and T.G. Brown

The Bercha Group
Suite 250, 1220 Kensington Rd. N.W.
Calgary, Alberta, Canada, T2N 3P5
Telephone: (403) 270-2221; Facsimile: (403) 270-2014; Telex: 03-827666

ABSTRACT

The concept of a fixed crossing linking Prince Edward Island and New Brunswick has raised concern over the associated socio-economic and environmental impacts. An important element of the environmental concerns involves ice climatological effects which could influence transportation, fisheries and agriculture through possible microclimatic changes.

The study described herein involved using simultaneously acquired C and X band synthetic aperture radar (SAR) data, acquired on February 19 and 20, 1988, for establishing baseline data on the ice conditions in the Northumberland Strait. These conditions included floe sizes, floe velocities, ridge frequencies and ice concentrations. In addition, imagery acquired using a real aperture radar (SLAR) system between December, 1987 and April, 1988, were used to provide a broader temporal distribution of the ice concentrations.

Keywords: SAR, SLAR, sea ice, ice monitoring

INTRODUCTION

A fixed crossing linking Prince Edward Island (P.E.I.) and New Brunswick (N.B.) has been the subject of debate since the 1960's. Recent plans to initiate the crossing have raised concern over the possible socio-economic and environmental impacts of the structure. An important element of the environmental concerns involves ice conditions which could influence vessel traffic and fisheries and even agriculture through possible changes in the microclimate. Therefore, establishing baseline data on the ice conditions within the Strait is the primary focus of this study. The baseline data on the ice conditions were divided into categories including floe sizes, ice dynamics, ridge frequencies and ice concentrations for input into an ice jam model. This mathematical model was used to determine the frequency and severity of ice jams induced by the fixed bridge crossing, and their effect on the length of the ice season.

STUDY AREA

The Northumberland Strait is located between the provinces of Prince Edward Island, Nova Scotia and New Brunswick (Figure 1). Its approximate length is 250 km and its width varies. The narrowest point of the Strait is 12 km and has been referred to as the Narrows. This point occurs between Borden, P.E.I., and Cape

Tormentine, N.B., and is the site of the proposed fixed link.

DATA

The sensor selected to acquire ice baseline data was the Canada Centre for Remote Sensing (CCRS) SAR. SAR has the advantages of producing high resolution imagery, regardless of weather and visibility conditions, and of acquiring images of large areas quickly. This project used SAR to obtain sequential C and X band images of the Northumberland Strait ice cover. A total of four swaths were acquired of the Strait area. Two swaths were flown on February 19 (Mission 1) between 23:34:41 and 00:36:07 Greenwich Mean Time (GMT), and two were obtained on February 20 (Mission 2) between 19:52:55 and 21:05:30 GMT. On each of the two days, an overlapping segment between the two swaths was acquired. The time difference between the overlapping areas was approximately 30 minutes. This overlap corresponded to the Narrows or the potential location of the fixed link. Mission 2 was flown approximately 20 hours after Mission 1. The area of coverage is shown in Figure 1.

These SAR flights simultaneously acquired C and X band data. Data records were in the form of real time X band analogue products, C and X band HDDT's (High Density

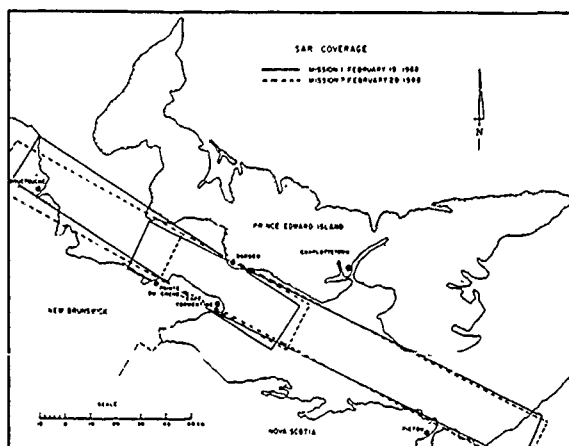


FIGURE 1: Study Area and SAR Coverage.

Digital Tapes), and C band VCR recordings. The HDDT's were processed into CCT's and positive and negative transparencies.

In addition to the SAR data, imagery from the Atmospheric Environment Service (AES) Motorola AN/APS 94D SLAR system was available. This imagery varied in scale from 1:500,000 to 1:1,000,000. A total of 40 AES SLAR transparencies were obtained for a period extending from December 28, 1987 to April 12, 1988.

Additional ancillary data acquired for the times during and between the SAR missions included wind conditions recorded by AES, Bedford, Nova Scotia, whether precipitation occurred, and water level information recorded by the Canadian Hydrographic Service at Charlottetown, P.E.I. and Pointe-du-Chêne, N.B.

METHODOLOGY

The SAR data were systematically analyzed to provide information on floe sizes, ice dynamics, ridge frequencies and ice concentrations. The SLAR data were analyzed to determine changes in conditions and determine ice motion information for the December through April time period. In addition, climate and water elevation data were acquired to correlate with the ice movement information obtained from the SLAR data.

FLOE SIZE

Floe sizes were measured using two procedures. First, all floes which intersected grid lines drawn in the range direction at 10 km intervals on the analogue imagery for the entire Strait were measured. In the regions 10 km wide on either side of the crossing line, the grid was reduced to 5 km. The second procedure involved measuring all floes on the real time analogue and enlarged digital images within an area of 5 x 18 km sampled at the Borden/Cape Tormentine Crossing. Both the longest axes (length) and the longest axes perpendicular to length (width) were measured. Any floe within the area or in contact with the boundary would be tallied and placed in a category based on dimension.

ICE DYNAMICS

Floe motion was calculated for various time periods using both SAR and SLAR data. Floes which were identifiable on successive images were selected for analysis. Essentially, the distance travelled and average time required to travel that distance, based on the floe centroid, was recorded and the approximate ice floe velocity calculated assuming motion along a vector. Two methods were used. The first involved short term dynamics and the second approach involved quantifying long term motion. In both cases, land masses were used to provide geographic orientation and registration of the information. The results of the ice motion were compared with wind and water level data for the same period. The long term motion methods were used to analyze the SLAR data. However, a smaller number of results were determined indicative of the difficulty in identifying floes from successive SLAR images.

ICE MORPHOLOGY

General ice conditions, including areas of open water, fast ice and ice concentrations, were mapped based on interpretations from SAR and SLAR data. Because of the small scale of the SLAR data, only a regional interpretation was possible.

C AND X BAND COMPARISON

A qualitative assessment of the C and X band images was made. Image aspects compared included tonal consistency, morphological definition, small object resolution, and internal feature definition.

RESULTS AND DISCUSSION

FLOE SIZES

The grid sampling approach resulted in a greater proportion of large floes in relation to smaller floes and ice cakes. This related to the fact that there was a greater probability of a larger floe intersecting the grid line. Therefore, the floe size frequency distributions were heavily weighted to large floes and resulted in a mean of 232 m. Because of the small scale of the analogue data (1:86,500), small floes and ice cakes were indistinguishable resulting in a large mean floe dimension.

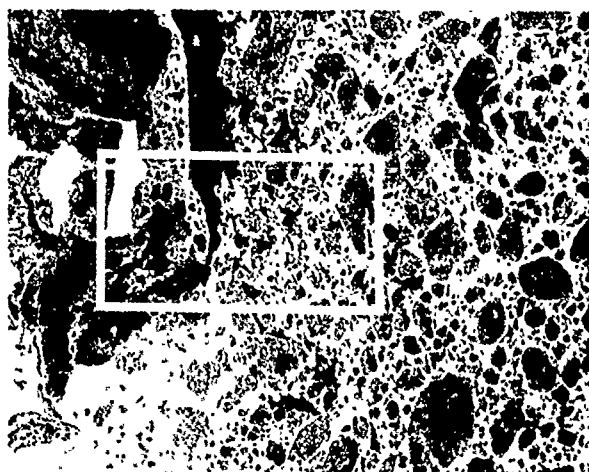
Using the areal sampling procedure, analogue (1:86,500) and digital X band images (1:33,000) were sampled (Figure 2). A greater percentage of small floes and ice cakes were recorded on the digital image than on the analogue image. As a result the mean floe dimension was 27 m as opposed to 116 m for the analogue data (Figure 3). With this method there is an equal probability of small and large floes occurring within an area and the resultant statistics were more reliable. These statistics, however, did not result in similar concentrations of ice cakes as were obtained by Brown et al (1989).

ICE DYNAMICS

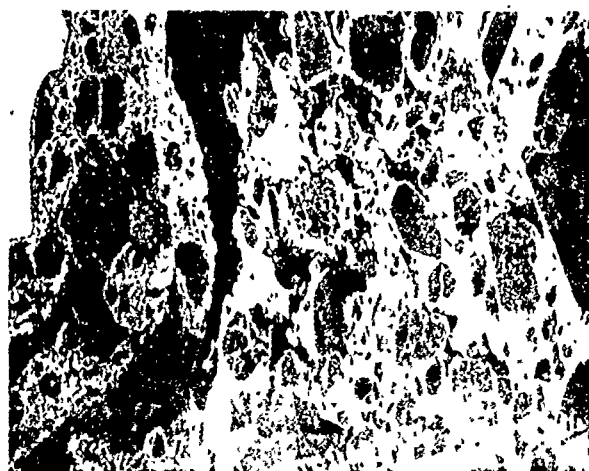
Short term dynamics were acquired by analyzing the overlapping segment at the crossing location for each mission. Floe motion for Mission 1 was from the northwest to the southeast. The results indicated an average floe velocity was 0.74 m/sec from 163 measured floes with the highest measured velocity, found in the crossing, of 1.14 m/sec. Velocity increased as the floes moved towards the narrows. This increase was associated with the constriction of floes within the Narrows where tidal streams have been recorded over 2 knots (Canada, 1985). Once through the Narrows, the velocity decreased which corresponded to the floes spreading out into a wider section of the Strait and similarly, the water currents being less restricted. At the time of the mission, water level data indicated an outgoing tide which accounts for the southeasterly motion.

In Mission 2, observed floe motion was from the southeast to northwest, opposite to the normal trend. The mean floe velocity was 0.61 m/sec calculated from 164 measurements. The highest velocity occurred in the crossing area at 1.13 m/sec (Figure 4). The combination of tidal flooding and strong southerly winds, 6 hours prior to the mission, resulted in a floe motion opposite to normal. Once the floes had passed through the Narrows, the velocity decreased in a fashion similar to Mission 1.

Floe motion between the two SAR missions was calculated to determine longer term dynamics. The general ice motion between the two missions was from southeast to northwest. The highest velocity occurred in the crossing area at 0.39 m/sec (Figure 5). From analysis of 191 floes for which floe movement was measured, a mean of 0.30 m/sec was calculated.



(a) Analogue SAR Image

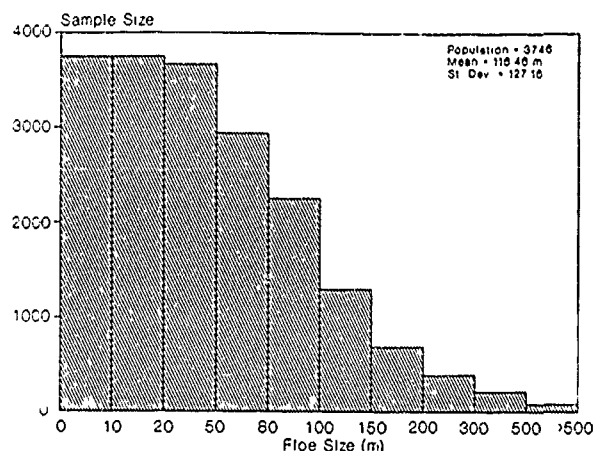


(b) Digital SAR Image

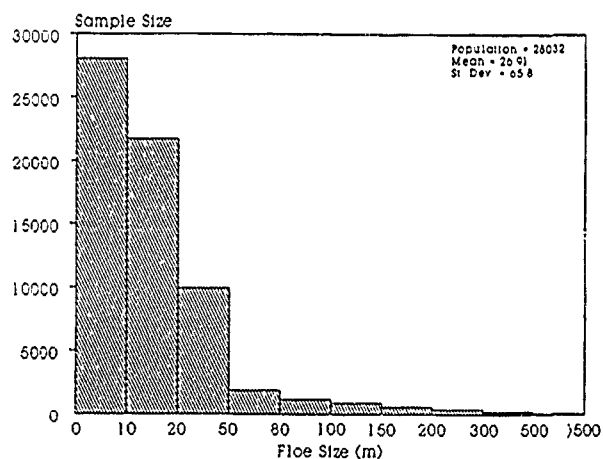
FIGURE 2: Comparison of SAR Image Products

When the results of long term and short term motion are compared, the difference between maximum velocities is conspicuous but not surprising (Table 1). As previously indicated, Mission 1 corresponded to the peak tidal flow, resulting in higher floe velocities. Similarly, the second mission was flown when high southeasterly winds preceded the flight by 6 hours. The resultant floe motion, therefore, is faster than the longer term motion and in a direction opposite to the norm. The effects of tides, currents and wind patterns on the floe velocity during the longer time interval were responsible for the lower velocities compared to short term velocities.

SLAR data provided an indication of floe dynamics throughout the months the data were collected. Floe motion varied in relation to the ice concentrations within the Strait, the position in the Strait (whether in a narrow or wide region of the Strait) and climatological and oceanographic effects between and during the times the images were acquired. Velocities varied from 0.02 m/sec to 0.48 m/sec (Table 1). This floe motion represents net velocities over a period of time (generally one or more days) and assumes a straight line motion.



(a) Cumulative Histogram of Analogue SAR Data



(b) Cumulative Histogram of Digital SAR Data

FIGURE 3: Floe Size Analysis

ICE RIDGING

Ice ridging was not evident from the analogue data source. Digital data revealed only minimal ridging. Field data obtained during the February time frame indicated ridging was scarce. Numerous ice accumulations were found along the edges of the floes, but these were determined to be rubble piles.

ICE MORPHOLOGY

By mapping the general ice conditions of the strait, it was possible to obtain an accurate picture of the development and disintegration of ice within the Strait over the winter and into the spring. In early December, open water occurred in the Strait west of the crossing with a high concentration of ice, 9/10ths, along the New Brunswick coast. As the ice season progressed, the concentration of ice increased throughout the Strait, with the exception of some areas along the landfast ice. The ice generally entered the Strait from the western end of P.E.I. and moved southeasterly. In March, large areas of open water developed, associated with ice breakup. By March 14 and 15, the ice began to

TABLE 1: Summary of Ice Motion Analysis Results

| | # OF OBSERVATIONS | MAX. VEL. (M/SEC.) | MEAN VEL. (M/SEC.) | STD. DEV. | RANGE | DIRECTION (AZIMUTH) |
|-------------------------------|----------------------|-----------------------|-----------------------|--------------|-----------------|------------------------|
| SHORT TERM MOTION (CLARS) | | | | | | |
| February 19 | 163 | 1.138 | 0.446 | 0.297 | 0.228- 1.138 | 120 |
| February 20 | 164 | 1.152 | 0.607 | 0.103 | 0.03- 1.152 | 289.5 |
| MEDIUM TERM MOTION (CLARS) | | | | | | |
| February 19-20 | 191 | 0.39 | 0.301 | 0.048 | 0.213- 0.39 | -- |
| LONG TERM MOTION (CLARS) | | | | | | |
| | 142 | 0.55 | 0.16 | -- | 0.01- 0.55 | -- |

**Range of several observations on different days

thin in the area west of the crossing and the higher concentrations moved eastward. Although most of the Strait was clear on March 25, the eastern end of the channel was still covered with 9-10/10ths ice.

C AND X BAND COMPARISON

The analysis was based on X band data because it had been assumed that the data contained superior information. Since C band data had been acquired, a qualitative comparison was made which suggested C band contained superior ice information content to X band.

First, the tonal characteristics of the C band imagery appeared more uniform, with better definition of morphological features. This may be attributed to lower attenuation due to the longer wavelength of C band as well as less local backscattering due to the significantly lower power associated with the C band. Also, internal ice floe features, such as ridging and rubble piles, could be better defined on the C band

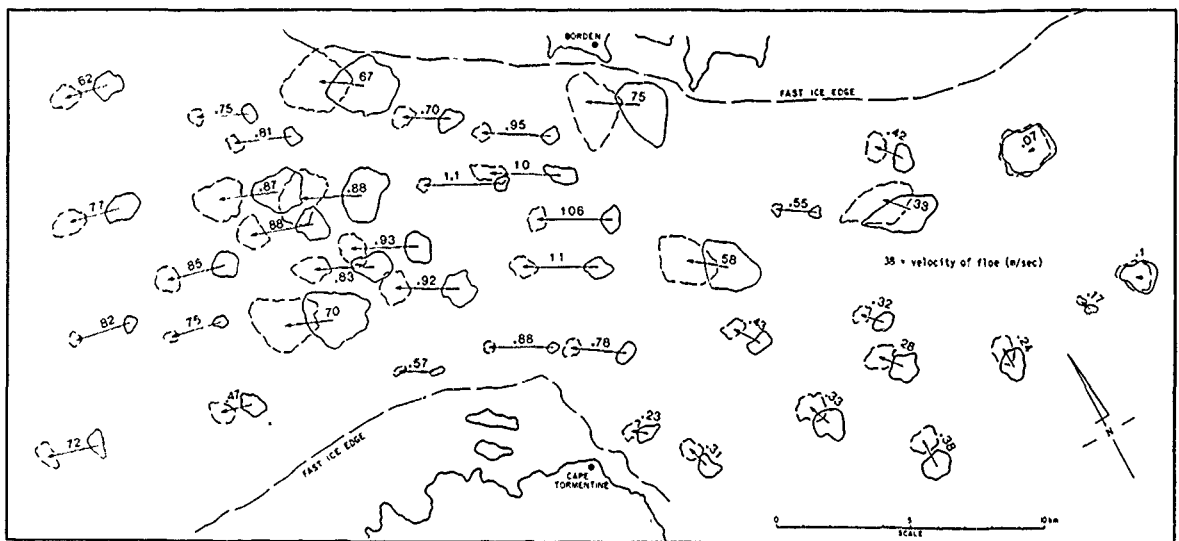


FIGURE 4: Short Term Floe Motion, Mission 2

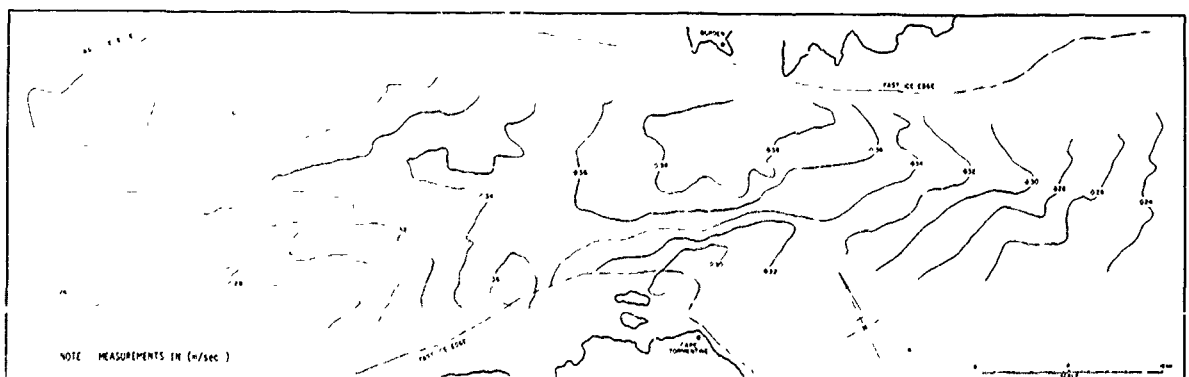


FIGURE 5: Floe Velocity Isoline Map, Mission 1 - Mission 2.

imagery than on the X band imagery. The latter may be attributable to the lower backscatter suggested by Askne and Johansson (1988).

CONCLUSIONS

The following conclusions can be made:

- a. SAR data provided an excellent source of information on floe sizes, ice conditions and ice movement. SLAR provided good regional coverage of ice conditions and ice movement. However, at the scale of the SLAR imagery, only the estimation of floe sizes greater than 150 meters is possible.
- b. Floe sizes obtained from the analogue data averaged 116 m. Digital data average floe sizes of 27 m.
- c. Real time analogue data provides good information on the kinematic and dynamic characteristics of the ice.
- d. The areal sampling technique should be selected over the grid approach in order to provide a more accurate representation of the distribution of floe sizes.
- e. Ice motion information can be obtained from both SAR and SLAR sequential imagery. Floes can be recognized on both types of imagery. In calculating ice motion, the approach assumes a constant vector floe motion, which is seldom the case in reality. In addition, ice motions should be correlated with climatological and oceanographic conditions to determine their effect on ice movement.
- f. Floe motion characteristics can be summarized as follows:
 - (i) Short term dynamics :
 - maximum velocity = 1.14 m/sec
 - mean velocity = 0.67 m/sec
 - (ii) SAR analysis (21 hours) :
 - maximum = 0.39 m/sec
 - mean = 0.30 m/sec
 - (iii) Long term SLAR analyses :
 - maximum = 0.480 m/sec
 - mean = 0.16 m/sec

The C and X band SAR has been shown to be an effective tool for the quantification of critical ice parameters. These data were used as data source for evaluating the impact of ice on a structure (Bercha, 1988).

ACKNOWLEDGEMENTS

Funding for the work described herein by Public Works Canada is hereby gratefully acknowledged.

The analysis described was directed by Valerie Shaw. Her untimely death, on February 23, 1989, prevented her from participating in the preparation of this paper.

REFERENCES

1. Askne, J., and Johansson, R., 1988, "Ice Ridge Observations by Means of SAR", Proceedings IGARSS 88, Edinburgh, Scotland.
2. Bercha, F.G. and Associates (Alberta) Ltd., 1988, "Northumberland Strait Crossing Ice Climate Study", Report to Public Works Canada, Hull, Quebec.
3. Brown, T.G., Dachka, J.A., and Steen, J.W., 1989, "Ice Kinematic Measurements for Aerial Photography", Proceedings IGARSS 89, Vancouver, B.C., Canada.
4. Canada, 1985, "Sailing Directions : Gulf and River St. Lawrence", Department of Fisheries and Oceans, Ottawa, 493 pp.

MEASURED ICE FLOE SIZES IN THE BEAUFORT SEA

William J. Stringer
 University of Alaska Fairbanks
 Fairbanks, Alaska, USA 99775-0800
 907/474-7455 (Telephone)
 907/474-7290 (Fax)
 35414 (Telex)

The areal extent of ice floes has been measured from Landsat imagery of the summertime Beaufort Sea, spanning the five months between break-up and freeze-up. Areas were measured by digitizing the floe boundaries directly from 1:50,000 scale photographic enlargements of Landsat images. After each floe field was digitized, the digitized boundaries were plotted by a mechanical plotter so that the digitized data could be compared with the original data set. Following this, each floe was numbered and its area computed by means of a raster-scanning computer program. The resulting floe size data were tested to determine whether they fit any numerical distribution. In general, the distribution of floe areas was found to obey a power law: $N(S) = N_1 S^\lambda$, where the counted number of floes per unit floe size interval, $N(S)$, is related to the number of floes in the particular distribution at unit floe size, (N_1), the floe size, (S), and λ , a parameter found here to range between -1.33 and -2.06. The value of λ decreased from -1.33 in May to -2.06 in August and then increased to nearly -1.47 in September. An exponential relationship with λ was found among the values of N_1 from the various distributions: $N_1 = N_0 e^{-15.14\lambda}$. This relationship appears to hold regardless of the seasonal variation of λ . Thus, floe size distributions were found to obey $N(S) = N_0 e^{-15.4\lambda} S^\lambda$, with a value of $N_0 = 1.23 \times 10^{-6}$, where N_0 is the projected number of floes per unit floe size at unit floe size for $\lambda = 0$.

Although not observed, a value of $\lambda = -1$ was found by theoretical considerations to produce a floe size distribution in which the apparent distribution of floe size is the same regardless of the scale at which it is viewed. Based on the observed variation of λ with season, it is hypothesized that such a distribution might appear earlier in the year than the observing period reported here.

ICE KINEMATIC MEASUREMENTS FROM AERIAL PHOTOGRAPHY

T.G. Brown, J.A. Dechka, J.W. Steen

The Bercha Group
Suite 250, 1220 Kensington Road N.W.
Calgary, Alberta, T2N 3P5
Phone: (403) 270-2221, Telefax (403) 270-2014; Telex: 03-827666

ABSTRACT

The Paper describes the analysis of an extensive photographic data base of the ice conditions in Northumberland Strait in 1964 and 1965. The original photographs were obtained from a series of overflights made throughout the two winters to monitor ice conditions in support of a previous development plan for a fixed crossing between Prince Edward Island and New Brunswick. The documentation available with the photographs was sufficient to permit scales to be developed for the photographs and hence measurements to be taken from the photographs.

Analysis of the photographic data resulted in considerable information on ice concentrations, ice floe sizes, and ridge frequencies. The ice floe size data indicated a very high concentration of very small floes or cakes, with the largest category in the 0 to 10 metre range. Ridge frequencies were found to vary significantly throughout the length of the Strait with concentrations varying from 15 per km to 50 per km.

The information obtained from the photographic data base was found to constitute a significant contribution to the available information on ice conditions and statistics in Northumberland Strait and was used in parallel analytical studies directed at determining the possible effects of various bridge configurations on the ice regime within the Strait. Because the photographs were acquired for a variety of operational, sensor, and target parameters, they required the development and application of some unique photogrammetric techniques. These included analysis of error bounds on the results.

KEYWORDS: Photographs, interpretation, ice, floes, ridges.

INTRODUCTION

The current proposal to construct a fixed link crossing between Prince Edward Island and New Brunswick across Northumberland Strait, has triggered significant concerns regarding the possible socio-economic and environmental impacts. One of the primary environmental concerns has been the impact of a fixed bridge crossing on the ice regime in Northumberland Strait and any modifications of that ice regime which might affect vessel traffic, the

fishing industry, and agriculture in the adjacent provinces. In response to these concerns, studies were carried out with the objective of characterizing the current ice regime within Northumberland Strait and determining any impacts which a fixed bridge crossing might have on this regime.

Clearly, both of the above objectives require the availability of a complete and reliable data base of ice conditions and ice kinematic statistics describing the ice regime within Northumberland Strait. The initial studies carried out to determine the effect of the bridge crossing on the ice regime and the effect of the ice regime on the bridge piers, identified serious deficiencies in the available data (Bercha and Associates, 1987; Acres, 1987). These deficiencies included information on floe sizes, ice thicknesses, ridge frequencies, and depths of consolidation. As a result of the gaps in the available ice data base, a number of projects were initiated by Public Works Canada to acquire the necessary data. The work described in this paper was based on a source of data which was available. In the mid-1960's, a previous proposal to construct a fixed link between Prince Edward Island and New Brunswick was carried to the point at which construction had commenced. In support of this development, a series of overflights surveying the entire extent of Northumberland Strait were carried out for two successive winters. These flights were used to obtain general information on the ice concentrations and extent of fast ice within the Strait, as well as to obtain a series of photographs of the ice conditions. Figure 1 shows the general region of Northumberland Strait with the individual flight lines identified.

The photographs were obtained with hand-held 35mm cameras and a variety of fixed focal length lenses. Fortunately, many of the photographs were supported by excellent documentation which provided aircraft altitude and camera configuration, as well as cross-reference to the location of the photograph. Preliminary examination of the photographs indicated that there was sufficient supporting documentation to permit scales to be deduced and hence dimensions in the horizontal plane to be determined. With this capability, the data could be used to obtain floe sizes and ridging frequencies, although other parameters, such as ridge height, could not.

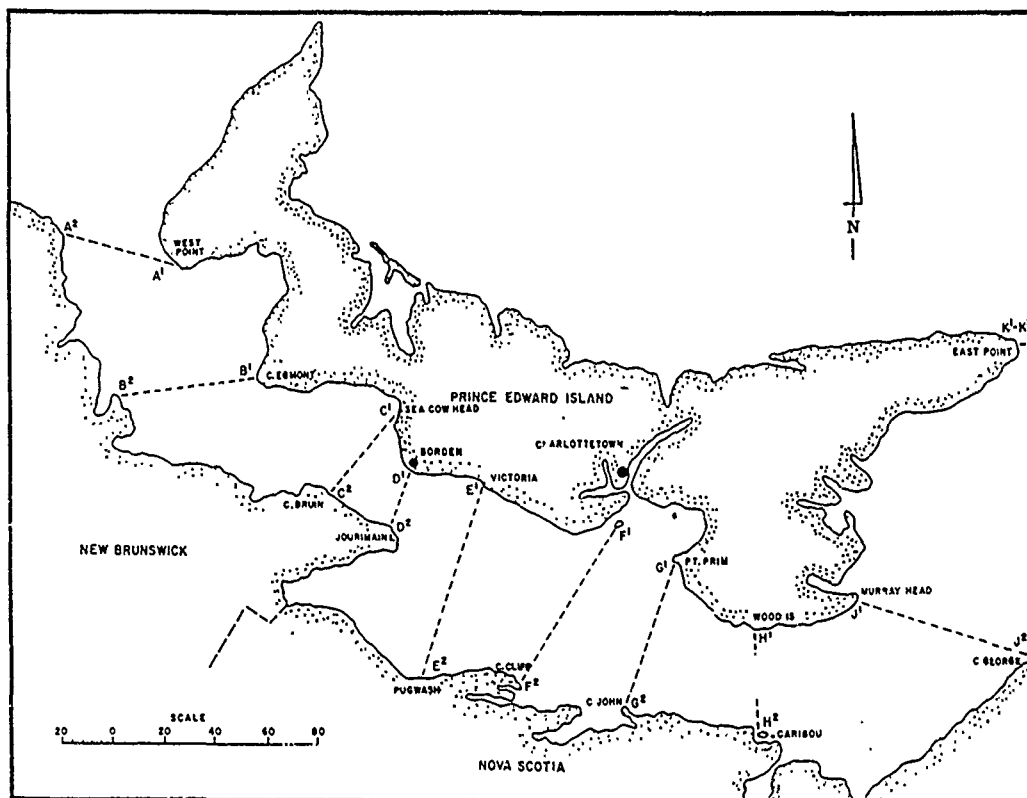


FIGURE 1 - NORTHUMBERLAND STRAIT AND FLIGHT LINES

DATA

The data, as received, consisted of approximately 1,000 photographs consisting of both colour slides and black and white prints. Negatives for the prints were available. The data base was reviewed to determine the availability of supporting documentation consisting of photograph location, aircraft altitude, and camera/lens configuration. Photographs for which all supporting documentation was not available, were not used for analysis. Ultimately, several hundred photographs were found to have sufficient documentation to permit analysis. Photographs of the fast ice regime at the two shorelines and photographs of open water were not included in the analysis of floe sizes although were clearly useful in the identification of general ice conditions.

Each of the selected photographs was reproduced to a fixed enlargement to simplify the interpretation process.

MEASUREMENT TECHNIQUES

The photographs could be divided into two distinct categories: high oblique photographs which included a horizon, and low oblique photographs which did not. The preparation of a perspective grid for a photograph for which the vanishing point can be determined, is well-defined in the literature (Estes et al, 1983).

Identification of the apparent horizon and the principal point of the photograph permits the calculation of the depression angle for that photograph. Once the depression angle is obtained, a grid with any previously specified grid spacing may be readily produced. Figure 2 gives an example of a high oblique photograph with a 100m grid overlain. The grids were developed by microcomputer and plotted on transparencies.



FIGURE 2 - HIGH OBLIQUE PHOTOGRAPH

The development of a grid for low oblique photographs is more difficult because no horizon is visible. For these photographs, the depression angle for the photograph must be estimated. Once this estimation has been made, the preparation of the photograph grid follows the same process as that used for the high oblique photograph. Nevertheless, the error in estimating the depression angle may lead to unacceptable errors in the determination of the perspective grid. Figure 3 shows the various angles involved in the development of the perspective grid for a low oblique photograph. In practice, it was found easier to estimate the angle α than the depression angle.

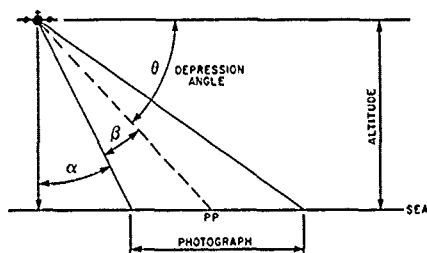


FIGURE 3 - LOW OBLIQUE ANGLES

For low values of α ($10-30^\circ$) the error in the grid scale induced by an error in the value of α was minimal. As the angle α became larger, the error induced in the grid scale increased to the point where such errors were unacceptable. Therefore, photographs that fell into this latter category were therefore not used for ice property measurement. Figure 4 gives an example of a low oblique photograph with the resulting grid superimposed. A number of photographs included, as subjects, identifiable objects whose dimensions were known. This allowed the estimation techniques for the angles to be refined.

RESULTS

For the floe size measurements, 10 floe size categories varying from 0 to 10m for the smallest category, to over 500m for the largest category were used. For each photograph, a grid was plotted on a transparency and overlaid on the photograph. Each identifiable floe was measured by first obtaining the largest lateral dimension and then obtaining the greatest dimension in a direction perpendicular to the first dimension. Floes were categorized according to their largest dimension. The number of floes measured on individual photographs varied widely, with some photographs yielding only a handful of ice floes while others yielding several hundreds. The data for each of the two years were kept separate as was data for different periods of each winter. This latter separation of the data was not entirely satisfactory as a number of periods of the winter were represented by only a few photographs. Early attempts to categorize the ice floe data by location within the strait were found to be unsupportable.

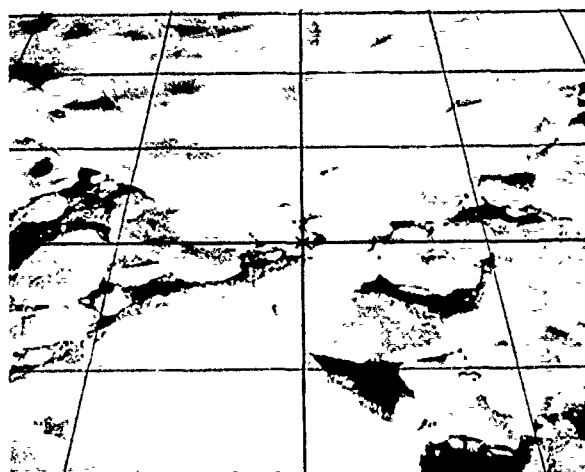
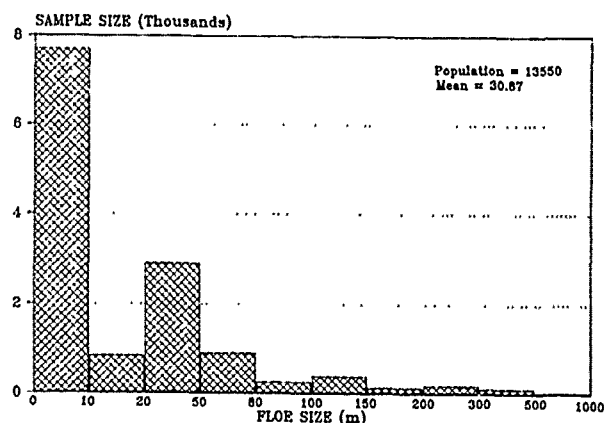


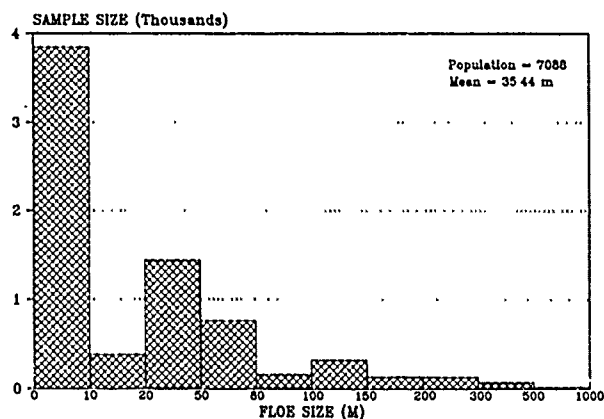
FIGURE 4 - LOW OBLIQUE PHOTOGRAPH

Figure 5 shows a histogram of ice floe data for the 1964 winter while Figure 6 presents a similar histogram for the 1965 winter. The heavy preponderance of floes in the 0-10m range is evident from both figures. Further, it may be noted that the

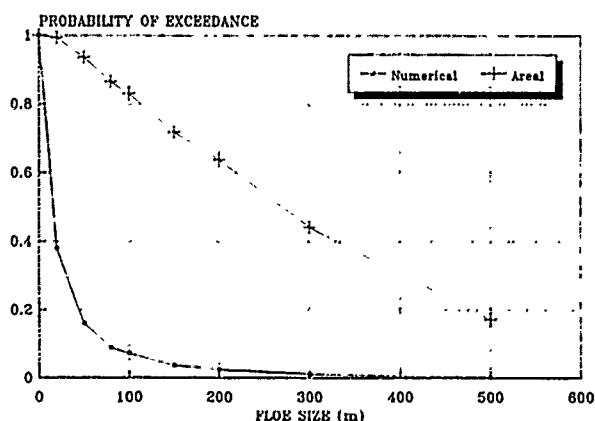
FIGURE 5: FLOE SIZE HISTOGRAM
1964



proportion of floes whose largest dimension is greater than 100m is quite small, less than 10% of the total. This is further illustrated in Figure 7 which shows the cumulative floe size distribution based on both years of data. This distribution is based on the data of Figures 5 and 6 and thus represents a numerical distribution curve for the floe size data. When the same data is presented as an areal concentration, the second (upper) curve of Figure 7 is obtained.

FIGURE 6: FLOE SIZE HISTOGRAM
1965

Some of the photographs were also used to obtain ridge frequency data. The frequency and the severity of the ridging process in Northumberland Strait was considered to be important in the evaluation of all ice-related effects of the construction of a bridge. While the measurement of ridge sail heights from the photographs was not possible, data was obtained on the frequency of ridging. This was carried out by obtaining greater enlargements of a number of photographs, preparing scaled grids similar to those prepared for the floe size data, and determining the frequency with which grid lines intersected ridges. The data resulting from the ridge frequency analysis is presented in Table 1 from which it may be seen that the average ridge frequency in Northumberland Strait lies between 20 and 30 ridges per kilometre with significant variations between the two years of data, between months of the winter, and even over regions of Northumberland Strait.

FIGURE 7 - FLOE SIZE PROBABILITY OF
EXCEEDANCE - 1964 & 1965TABLE 1 - RIDGE FREQUENCY RESULTS
(Ridges/km)

| PERIOD | LOCATION | | | |
|------------|-----------------|---------------------------------------|---|---------------------------------------|
| | Complete Strait | West of Line Cap Bruin - Sea Cow Head | Cap Bruin - Sea Cow Head to Borden-Tormentine | Borden-Tormentine to Wool Is.-Caribou |
| February | 33.7 | 21.3 | 45.8* | 42.7 |
| March | 22.6 | 22.4 | 20.4 | 24.7 |
| April | 25.2 | - | - | 25.6 |
| March 1964 | 25.2 | | | |
| March 1965 | 19.7 | | | |

*Limited data

CONCLUSIONS

In conclusion, an archived data base of photographs of ice conditions in a specific region within the Gulf of St. Lawrence represented a valuable source of information on ice properties in that area. Using existing photogrammetric techniques and some flexibility in interpretation, the data base yielded a significant contribution to the available information on floe size distributions and ridge frequencies within Northumberland Strait. It was found that the numerical concentration of ice floes was significantly biased towards the small floe size of cake in the category 0-10m; however, when presented as aerial concentrations, the probability of exceedance curve became almost linear. Ridge frequencies within Northumberland Strait varied between 20 and 30 ridges per kilometre with significant variations between different areas of the Strait.

ACKNOWLEDGEMENT

This paper is presented in the memory of Valerie L. Shaw whose enthusiasm and guidance contributed significantly to the work described.

The authors acknowledge the support of Public Works Canada and, in particular, the support of Mr. J. Feltham.

REFERENCES

1. Acres International Ltd., 1987, "Northumberland Strait Bridge Ice Forces Report", Report to Public Works Canada, March.
2. Bercha and Associates, 1987, "Northumberland Strait Crossing, Ice Climate Study", Report to Public Works Canada, October.
3. Bercha and Associates, 1988a, Addendum Report, "Northumberland Strait Crossing Ice Climate Study", Report to Public Works Canada, June.
4. Bercha and Associates, 1988b, "Impact of Bridge Proposals on Northumberland Strait Ice Regime", Report to Public Works Canada, November.
5. Estes, J.E. et al., 1983, "Fundamentals of Image Analysis; Analysis of Visible and Thermal Infrared Data" in Colwell, ed., Manual of Remote Sensing, pp 987-1124.

THE DETERMINATION OF ICE DISPLACEMENTS FROM SEQUENTIAL SAR IMAGERY

K.D. Oliphant, J.F. Sykes, and E.D. Soulis

Department of Civil Engineering, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada

Tel:(519)885-1211 Telex:069-55259 Fax:(519)888-6197

Abstract— Although SAR imagery is recognized as the best medium for providing ice motion information, it is not free from the possibility of error. Problems arise due to the vastness of the image data set, the presence of noise, the lack of ground control points and the idiosyncrasies of the material behavior of ice as they affect remote sensing. A set of 191 ice displacements was extracted from two sequential SAR images and the error present in these displacements was assessed.

Keywords—SAR imagery, pack ice motion, Lancaster Sound

1 Introduction

SAR imagery of Arctic pack ice has become the major source of ice behavior information because of its ability to provide 24 hour per day, all weather coverage (Vesecky et al, 1986). SAR also has the capability to provide data at a dense sampling rate (Hall and Rothrock, 1981); this is of importance to real-time forecasting (Fily and Rothrock, 1986).

However, SAR imagery can be far from the ideal medium to relay ice pack information. The size of an image data set can be unmanageably large, both in terms of the computer disk space required to store it and the active memory required during manipulation. Further, the bulk of the data set is usually unjustified as SAR imagery provides far higher resolution than is necessary for normal uses. Also, speckle or noise can severely degrade the quality of SAR imagery and SAR images are often difficult to register accurately to a coordinate system since they may contain few, if any, ground control points. If land features are present in the imagery, they can be used to register the images (as can maps), but distortion within the imagery can provide a significant problem. The current technology of the inertial navigation systems which are an alternative source of ground location information is also poor. Other problems arise as a result of the material characteristics of the ice which is remotely sensed. The SAR senses the electrolytic properties of the surface below it; this means that the signal reflected back is affected by the salinity and temperature of the ice (Curran, 1985), the presence of snow (Kim et al, 1985) and seasonal thaw (Livingstone et al, 1987). This study determines ice displacements from sequential SAR imagery and gives an expression for the maximum possible error inherent in the imagery and displacements.

2 SAR Image Data Set

The data set for this study consisted of a pair of images (Figures 1 and 2) from the 1986 STAR 1 reconnaissance flights by CAN-ARCTIC (de Bastiani, 1987). The imagery was geometrically rectified by Intera Technologies of Calgary, ALTA.

The images show the ice floe activity in Lancaster Sound near Brodeur Peninsula on Dec. 5th and 9th, 1986. The ice is moving predominantly east (de Bastiani, 1987). Large floes from Barrow Strait are responsible for the rafting of long, thin floes of ice (finger rafting). A shear plane exists at the mouth of Navy Board Inlet due to last fast ice in the inlet.

3 Image Preprocessing

3.1 Data Compression

Before compression, common resolutions of SAR images are 25 m and 4 or 12 m for satellite and aircraft originating images respectively. This high resolution leads to problems when storing and manipulating such an excessive data set. In order to contain the data for the motion in an area of several hundred kilometres, thousands of megabytes (MB) of disk space may be required. Graphic display packages and image processing systems may also be restricted in the amount of data they can manage. Since ice analysts recognize 100 m pixel resolution as sufficient to describe the ice motion (DeAbreu, 1987), the size of the data set could be greatly reduced if the resolution was lower. This study compressed the original images by a factor of 8 in both directions using a median filter; that is, each 8×8 pixel area was replaced by the median of the area. The resulting size of the images was 512×512 pixels or 50×50 km which implies a pixel resolution of approximately 100×100 m.

3.2 Speckle

The dominant radiometric noise source in SAR is speckle (Hudson, 1987). It has been shown to be signal dependent by modeling it as multiplicative noise (Curlander et al, 1985). However, despite efforts to extract information from speckle, it generally degrades the quality of imagery. Therefore, it is desirable to eliminate speckle with minimum loss of resolution. Speckle is normally removed using filters which consist of replacing a pixel window by the mean, median or mode value of the area. Unless the im-

ages are very poor quality (eg. fuzzy), the speckle content of an image is usually insignificant after compression. In this study, speckle was significantly reduced by the 8×8 pixel median filter compression and so no further processing was necessary.

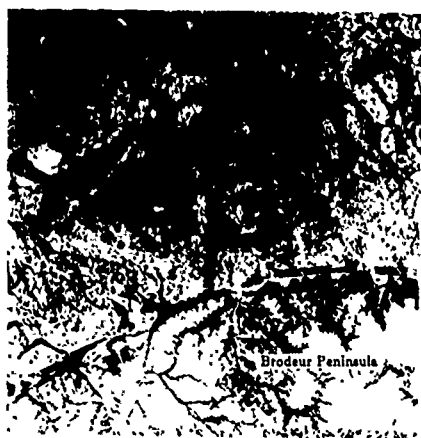


Figure 1: December 5, 1986 radar image



Figure 2: December 9, 1986 radar image

4 Image Registration

The generation of observed displacements from the sequential images requires the registration of the imagery. The images were first mapped together using a set of tie points of easily identifiable land features and the coordinates were calculated relative to the southwestern corner of the Dec. 5th image. Because of the orientation of the land in the images, the tie points were well distributed in only one of the cartesian directions with ten points being selected. The image offset was an average of 441.0 and -5.9 pixels (approximately 44.1 km and -0.6 km) in the azimuth (x) and range (y) directions respectively.

To test the difference between the ten data pairs and their mean offset, a trend representing a discrete difference between the two images as opposed to random distortion, was investigated. The distortion in both the azimuth and range directions was estimated. The range distortion check was limited to the width of the land (about $\frac{1}{5}$ of the width of the image). A regression between the range offset Δy and the location (x, y) was made and a standard statistical test at 95% confidence found Δy

to be constant over all (x, y), that is

$$\Delta y_{dec9-dec5} = C_0 \quad (1)$$

where C_0 is -5.9 pixels.

For the azimuth distortion check, points could be taken along the complete side since land was present along an entire side of the imagery. A regression of the azimuth offset Δx and the location (x, y) and a 95% confidence limit did detect a dependence of Δx with x . It can be written as

$$\Delta x_{dec9-dec5} = B_0 + B_1 x \quad (2)$$

where $B_0 = 434$ pixels and $B_1 = 0.016$. This dependency could be the result of a constant acceleration by the aircraft on either or both of the Dec. 5th or Dec. 9th flights. For x ranging from 0 to 512 pixels, the maximum value of $B_1 x$ is 8 pixels. This error was compensated for by subtracting off the distortion from each data point as a function of x .

5 Extracting Ice Displacements

Homologous ice features were selected using a digitizing program and the relative change in position of a floe from Dec. 5th to Dec. 9th gave the displacement. In total, 191 pairs of points were selected (Figure 3). In agreement with the assessment of the ice motion in Lancaster Sound during Dec. 1986 by de Bastiani (1987), the floes were found to be strongly eastward and linear. Considering the effective ice pack area in the images (35×50 km), the density of data points was equivalent to approximately 0.11 data point per km^2 . This density is comparable with that obtained by Hall and Rothrock (1981). For their 20 km scale ice motion study, they used a density of $\frac{1 \text{ pt}}{2 \text{ km}}$ or $\frac{0.25 \text{ pt}}{km^2}$. This study observed a larger scale of ice motion (on a scale of 44 km) and so a $0.11 \frac{\text{pt}}{km^2}$ density is reasonable.

There are four contributing sources of error in the observed ice displacements; they occur as a result of errors in cursor positioning, feature detection, image resolution and distortion in both the azimuth and range directions, i.e. along and across the track.

The error to pinpoint the cursor is a systematic error which represents the difference between the selected point and the point which the cursor actually digitizes. It is quantified as ± 100 m for both points of the displacement pair. The second source of error, the error in feature detection, consists of a ± 100 m error for each point of the displacement pair to allow for the possibility that the selected feature lies right on a pixel boundary. There is also a minimum error of ± 100 m for the second point in order to ensure that it represents the same feature as the first point. If the images or sections of it are blurry, then this minimum value will increase.

The resolution of the image is the third contributing source of error. As a result of the resolution, features may appear up to a pixel away from where they actually are and no displacements can be measured which are less than the resolution of the system (Curlander et al, 1965). In the present study the pixel, or elemental image unit, represented $100 \text{ m} \times 100 \text{ m}$ of actual coverage; this leads to a possible error of:

$$\begin{aligned} \epsilon^2 &= \epsilon_x^2 + \epsilon_y^2 \\ &= 100^2 + 100^2 \\ &= 20000 \text{ m}^2 \end{aligned} \quad (3)$$

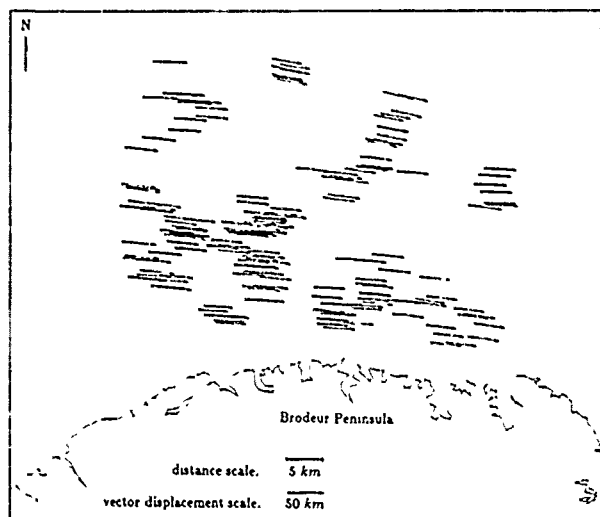


Figure 3: Observed Displacement Data Set

so that $\epsilon = 100\sqrt{2}$ or about 140 m in each axial direction. However, this 140 m resolution error is only a lower bound. The actual minimum size of the floe which can be identified depends on the features present in the imagery. Curlander et al (1985) state that in areas of uniform backscatter, floes had to be at least 80 pixels long before they could be identified.

The final contribution to the error in the observed ice displacements is due to the error in registering the images to one another. Although the distortion component could be identified and removed, the random error component (or unexplained variation) resulting from image registration must be included in the sources of error. The random error as determined by the regression of equations (1) and (2) is 110 m and 140 m in the azimuth and range directions respectively.

Table 1. Components of error for observed displacements

| Error Component | 1st Point of Pair | 2nd Point of Pair | Total For Pair |
|--------------------|-------------------|-------------------|----------------|
| Cursor Positioning | ± 100 m | ± 100 m | ± 200 m |
| Feature Detection | ± 100 m | ± 200 m | ± 300 m |
| Resolution | ± 140 m | ± 140 m | ± 280 m |
| Registration | ± 110 m | ± 140 m | ± 250 m |
| Total Error | ± 450 m | ± 580 m | ± 1030 m |

The magnitudes of the individual error components are tabulated in Table 1. While the errors estimated for each of the four components discussed in the preceding paragraphs may have some cancelling effect, the errors are added in this study to determine the maximum possible error for the ice displacements. From Table 1 the total error for each displacement is ± 1030 m. Compared to the average observed displacement of 44 km, the relative error for a single displacement is

$$\frac{1030 \text{ m}}{44 \text{ km}} = 2.3 \% \quad (4)$$

In a further exercise a statistical method of vector interpolation called co-kriging was used for the estimation because of its ability to estimate the original data points as well as provide the variance of the estimates (McKenna, 1986). Both of these features are useful in evaluating the accuracy of the interpolator.

Co-kriging was used to estimate the displacement values at regular intervals of approximately every 7 km. This spacing was compatible with the specifications required by ice analysts and navigators in ice motion forecasting (DeAbreu, 1987). The co-kriging procedure was performed using all the (191) observed displacement points. The vector field is shown in Figure 4.

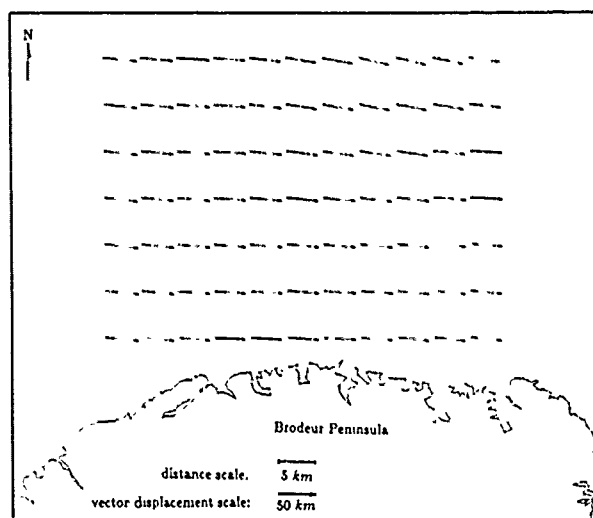


Figure 4: Co-kringed Displacement Data Set

6 Conclusions

After compressing the data set to a manageable size and reducing speckle, but still maintaining a pixel resolution acceptable to ice analysts, the images were registered relative to each other. A linear distortion which varied from 0 to 8 pixels along the azimuth dimension was detected with 95% confidence. It was caused by a surge in the aircraft's velocity in the flight path direction. Distortion in the range direction was not significantly different from zero for the range data which could be extracted. The total error in the displacements was given as the sum of the components despite the fact that some may cancel. The maximum possible error in the 191 ice displacements was ± 1030 m of 2.3% relative to the average displacement.

7 Acknowledgement

Funding for this study was provided by the National Science and Engineering Research Council (NSERC) post graduate scholarship awarded the senior author and the NSERC strategic grant entitled: *The Image Analysis of Sea Ice*.

References

- [1] de Bastiani, P., "Canadian Arctic Marine Ice Atlas: Winter 1986-1987", Transport Canada Report No. TP-8318-E, 66p, 1987.
- [2] Curlander, J.C., B. Holt and K.J. Hussey, "Determination of Sea Ice Motion Using Digital SAR Imagery", IEEE J. of Oceanic Eng., Vol. OE-10, No. 4. pp358-367, 1985.

- [3] Curran, P.J., "Principles of Remote Sensing", Longman Group Limited, London, 1985.
- [4] DeAbreu, R., "Four and Eight Bit SAR Image Reduction: A Qualitative Analysis", 3B Work Report for the Dept. of Geography, University of Waterloo, Waterloo, Ontario, 24p, 1987.
- [5] Fily, M. and D.A. Rothrock, "Extracting Sea Ice Data from Satellite SAR Imagery", IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-24, No. 6, 1986.
- [6] Hall, R.T. and D.A. Rothrock, "Sea Ice Displacement From SEASAT Synthetic Aperture Radar", J. of Geophysical Res., Vol. 86, No. C11, pp11078-11082, 1981.
- [7] Hudson, D.V., "SAR Imaging and Representation of Sea Ice", GEOG675 Project, Dept. of Geography, Fac. of Env. Studies, University of Waterloo, Waterloo, Ontario, 41p, 1987.
- [8] Kim, Y.S., R.K. Moore, R.G. Onstott and S. Gogineni, "Towards Identification of Optimum Radar Parameters for Sea-Ice Monitoring", J. of Glac., Vol. 31, No. 109, pp214-219, 1985.
- [9] Livingstone, C.E., R.G. Onstott, L.D. Arsenault, A.L. Gray and K.P. Singh, "Microwave Sea-Ice Signatures Near the Onset of Melt", IEEE Trans. on Geoscience and Remote Sensing, Vol. GE-25, No. 2, pp174-187, 1987.
- [10] McKenna, R.F., "Spatial Analysis of Ice Pack Motion", Ph.D. Thesis, Dept. of Civil Engineering, University of Waterloo, Waterloo, Ontario, 189p, 1986.
- [11] Vesecky, J.F., R. Samadani, M.P. Smith, J.M. Daida and R.N. Bracewell, "Automated Remote Sensing of Sea Ice Using Synthetic Aperture Radar", Proc. of IGARSS' 86 Symposium, Zurich, Switzerland, Ref. ESA SP-254, 1986.